NATIONAL CENTER FOR EDUCATION STATISTICS

NAEP 1996 SCIENCE

Report Card for the Nation and the States

THE NATION'S REPORT CARD naep

U.S. DEPARTMENT OF EDUCATION OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

Findings from the National Assessment of Educational Progress

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history/geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics, the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations. NAEP reports directly to the Commissioner, who is also responsible for providing continuing reviews, including validation studies and solicitation of public comment, on NAEP's conduct and usefulness.

In 1988, Congress established the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The Board is responsible for selecting the subject areas to be assessed from among those included in the National Education Goals; for setting appropriate student performance levels; for developing assessment objectives and test specifications through a national consensus approach; for designing the assessment methodology; for developing guidelines for reporting and disseminating NAEP results; for developing standards and procedures for interstate, regional, and national comparisons; for determining the appropriateness of test items and ensuring they are free from bias; and for taking actions to improve the form and use of the National Assessment.

The National Assessment Governing Board

NATIONAL CENTER FOR EDUCATION STATISTICS

NAEP 1996 Science Report Card for the Nation and the States

Findings from the National Assessment of Educational Progress

Christine Y. O'Sullivan
Clyde M. Reese
John Mazzeo

May 1997

Office of Educational Research and Improvement
U.S. Department of Education

Prepared by Educational Testing Service under a cooperative agreement
with the National Center for Education Statistics

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Ramon C. Cortines
Acting Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

Education Assessment Group
Gary W. Phillips
Associate Commissioner

May 1997

SUGGESTED CITATION
O'Sullivan, C.Y., Reese, C.M., and Mazzeo, J., *NAEP 1996 Science Report Card for the Nation and the States,* Washington, DC: National Center for Education Statistics, 1997.

FOR MORE INFORMATION
Contact:
Arnold A. Goldstein
202-219-1741

For ordering information on this report, write:

National Library of Education
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, D.C. 20208-5641

or call 1-800-424-1616 (in the Washington, DC, metropolitan area call 202-219-1651).

This report also is available on the World Wide Web: http://www.ed.gov/NCES/naep.

The work upon which this publication is based was performed for the National Center for Education Statistics, Office of Educational Research and Improvement, by Educational Testing Service.

Table of Contents

TABLES

FIGURES

Executive Summary

NAEP 1996 Science Report Card for the Nation and the States[1]

For more than a quarter of a century, the National Assessment of Educational Progress (NAEP) has reported to policy makers, educators, and the general public on the educational achievement of students in the United States. As the nation's only ongoing survey of students' educational progress, NAEP has become an important resource for obtaining information on what students know and can do.

The NAEP 1996 science assessment continues the mandate to evaluate and report the educational progress of students at grades 4, 8, and 12. The national results provided herein describe students' science achievement at each grade and within various subgroups of the general population. State-level results for grade 8 are presented for the 44 individual states and other jurisdictions that chose to participate in the 1996 state assessment and met the guidelines for participation. NAEP national and state data assess the performance of students in both public and nonpublic schools.

The NAEP 1996 Science Framework

The science assessment was crafted to measure the content and skills specifications described in the science framework for the 1996 National Assessment of Educational Progress. Two organizing concepts underlie the science framework. First, according to the framework, scientific knowledge should be structured so as to make factual information meaningful. The way in which knowledge is structured should be influenced by the context in which the knowledge is being presented. Second, science performance depends on knowledge of facts, the ability to integrate this knowledge into larger constructs, and the capacity to use the tools, procedures, and reasoning processes of science to develop an increased understanding of the natural world. Thus, the framework called for the NAEP 1996 science assessment to include the following:

- Multiple-choice questions that assess students' knowledge of important facts and concepts and that probe their analytical reasoning skills;
- Constructed-response questions that explore students' abilities to explain, integrate, apply, reason about, plan, design, evaluate, and communicate scientific information; and
- Hands-on tasks that probe students' abilities to use materials to make observations, perform investigations, evaluate experimental results, and apply problem-solving skills.

[1] The Executive Summary for this report was prepared by Alan Vanneman of the Education Statistics Services Institute.

The core of the science framework is organized along two dimensions. The first dimension divides science into three major fields: earth, physical, and life. The second dimension defines characteristic elements of knowing and doing science: conceptual understanding, scientific investigation, and practical reasoning. Each question in the assessment is categorized as measuring one of the elements of knowing and doing within one of the fields of science (e.g., scientific investigation in the context of earth science). The framework also contains two overarching domains — the nature of science and the organizing themes of science. The nature of science encompasses the historical development of science and technology, the habits of mind that characterize science, and the methods of inquiry and problem solving. It also includes the nature of technology — specifically, design issues involving the application of science to real-world problems and associated trade-offs or compromises. The themes of science include the notions of systems and their application in the scientific disciplines, models and their functioning in the development of scientific understanding, and patterns of change as they are exemplified in natural phenomena.

Student Achievement

Students' science performance is summarized on the NAEP science scales, which range from 0 to 300 at each grade. While the scale-score ranges are identical, the scales were derived independently at each grade. Therefore, average scale scores across grades cannot be compared. For example, equal scale scores on the grade 4 and grade 8 scales do not imply equal levels of science achievement. Within each of the three grades, scale scores for students ranged from about 105 for those scoring at the 10th percentile to about 192 for those performing at the 90th percentile.

It is possible to illustrate the level of achievement of students with a given scale score by identifying questions likely to be answered correctly by students with that scale score, a process known as "mapping." The position of the question on the scale represents the scale score attained by students who had at least a 65 percent probability of reaching a given score level on a constructed-response question or at least a 74 percent probability of correctly answering a multiple-choice question. Mapping questions onto the NAEP science scales can be used to illustrate the range of achievement of students *at or near* selected percentiles. For example, eighth graders at or near the 50th percentile were likely to correctly identify the effect of acid rain. Put slightly differently, this question was answered correctly by at least 74 of every 100 students scoring at or above the 150 scale-score level.

Sample Questions At or Near Selected Percentiles

THE NATION'S REPORT CARD naep

Grade 4	
10th	Identify items that conduct electricity. (105)
25th	Read the level of a liquid in a graduated cylinder. (129)
50th	Infer the function of animal teeth from diagrams showing their structure. (152)
75th	Explain the impact of fish death on an ecosystem. (173)
90th	Explain why Earth never runs out of water. (192)
Grade 8	
10th	Find typical yearly rainfall from a graph. (104)
25th	Explain the impact of fish death on an ecosystem. (127)
50th	Identify the effect of acid rain. (150)
75th	Understand the location of earthquakes. (172)
90th	Explain why lightning is seen before thunder is heard. (194)
Grade 12	
10th	Determine which planet has the longest year given data. (99)
25th	Describe how to avoid electric shocks in the home. (125)
50th	Identify what happens when a magnet is placed inside a coil. (150)
75th	Understand which evidence supports continental drift theory. (177)
90th	Understand structure and function of cell parts. (197)

The value in parentheses represents the scale score attained by students who had a 65 percent probability of reaching a given level on a constructed-response question or a 74 percent probability of correctly answering a 4-option multiple-choice question.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Major Findings for the Regions and States[2]

NAEP data can be used to compare student performance of fourth-, eighth-, and twelfth-grade students attending schools in four different regions — Northeast, Southeast, Central and West — and state-level results for eighth-grade students.

- For all three grades, students in the Central region had higher average scores than students in the Southeast region.
- In 19 of the 44 participating jurisdictions,[3] the average scale score for public school eighth graders was higher than the national average, while 14 jurisdictions performed below this average. The remaining 11 jurisdictions performed at or around the national average.

Summary of Jurisdiction Performance Relative to the Nation for Grade 8 Public Schools

THE NATION'S REPORT CARD naep

Performed Above the National Average	Performed At or Around the National Average	Performed Below the National Average
Alaska ‡	Arizona	Alabama
Colorado	Kentucky	Arkansas ‡
Connecticut	Maryland ‡	California
DDESS	Missouri	Delaware
DoDDS	New York ‡	District of Columbia
Indiana	North Carolina	Florida
Iowa ‡	Rhode Island	Georgia
Maine	Texas	Guam
Massachusetts	Virginia	Hawaii
Michigan ‡	Washington	Louisiana
Minnesota	West Virginia	Mississippi
Montana ‡		New Mexico
Nebraska		South Carolina ‡
North Dakota		Tennessee
Oregon		
Utah		
Vermont ‡		
Wisconsin ‡		
Wyoming		

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public school participation rates (see Appendix A).

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools

DoDDS: Department of Defense Dependents Schools (Overseas)

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

[2] In all discussions of differences in academic performance between subgroups, only statistically significant differences are reported. Such differences are unlikely to be due to chance factors.

[3] Several states participated but failed to meet established participation guidelines for reporting results. See Appendix A for more complete information of jurisdictions' participation rates.

Major Findings for Student Subgroups

The NAEP 1996 science assessment reports national data on the basis of demographic subgroups, level of parental education, type of school attended, and participation in selected government programs.

- Male and female students in grades 4 and 8 had similar scores. However in grade 12, male students had higher scores than female students.
- White and Asian/Pacific Islander students had higher average scores than Black and Hispanic students at all three grades.
- American Indian students had higher average scores than Black and Hispanic students, in grades 4 and 8. (The sample of American Indian students at grade 12 was too small to permit comparisons.)
- In general, at all three grades higher levels of parental education were associated with higher levels of student performance.
- At all three grades, students who attended nonpublic schools had higher average scores than those who attended public schools.
- Fourth- and eighth-grade students receiving services supplied by Title I programs had lower scale scores than those who did not participate in Title I. (The sample for twelfth graders who participated was not large enough to permit a comparison.) Title I of the Elementary and Secondary Education Act provides funding to local educational agencies to meet the needs of children who are economically disadvantaged and who are performing below grade level.
- At all three grades, students eligible for the free/reduced-price lunch program administered by the U.S. Department of Agriculture (USDA) scored lower than those not eligible. Eligibility for free/reduced-price lunches is determined by the USDA's Income Eligibility Guidelines. (Information on eligibility was lacking for 12 percent of fourth graders, 23 percent of eighth graders, and 21 percent of twelfth graders.)

Exploring a More Inclusive NAEP

An area in which the NAEP program continues to seek improvements is in the inclusion and appropriate assessment of two specific populations: students with disabilities (SD) and limited English proficient (LEP) students. The NAEP 1996 math and science assessments included supplemental samples of schools and students to enable the program to study the effects of revised inclusion rules on assessment results and to investigate the feasibility and impact of increasing the participation of students with disabilities and LEP students by offering assessment accommodations and adaptations.

Results from the grade 8 state NAEP science assessment indicated that the use of revised inclusion criteria, without the provision of accommodations, had little effect on the overall percentage of the total population assessed, or on the percentages of students with disabilities or LEP students assessed. There was some evidence from the national NAEP assessment that the provision of accommodations resulted in higher rates of participation for both groups of students.

Chapter 1

NAEP 1996 Science Assessment

NAEP's Mission

The National Assessment of Educational Progress (NAEP) is the only nationally representative and continuing assessment of what students in the United States know and can do in various academic subjects. NAEP is authorized by Congress and directed by the National Center for Education Statistics (NCES) of the U. S. Department of Education. The National Assessment Governing Board (NAGB), an independent body, provides policy guidance for NAEP.

Since its inception in 1969, NAEP's mission has been to collect, analyze, and produce valid and reliable information about the academic performance of students in the United States in various subject areas. In 1990, the mission of NAEP was expanded to include state-by-state results. State participation in NAEP is voluntary and has grown from 40 states and territories in 1990 to 47 in the 1996 science assessment. NAEP has also become a valuable tool in tracking progress toward the National Education Goals. The subjects assessed by NAEP are those highlighted at the 1989 Education Summit and in later legislation.[1]

The primary purpose of this report is to inform policy makers and the public about student achievement in science.

The NAEP 1996 Science Framework

The science assessment was crafted to measure the content and skills specifications described in the science framework for the 1996 National Assessment of Educational Progress.[2] The framework was developed in 1991 through a consensus process involving educators, policy makers, science teachers, representatives of the business community, assessment and curriculum experts, and members of the public. The project was managed by the Council of Chief State School Officers (CCSSO) under the auspices of NAGB.

---

[1] Executive Office of the President, *National Goals for Education* (Washington, DC: Government Printing Office, 1990); Goals 2000: Educate America Act, P.L. 103-227 (1994).

[2] *Science Framework for the 1996 National Assessment of Educational Progress* (Washington, DC: National Assessment Governing Board, 1995).

Two organizing concepts underlie the science framework. First, according to the framework, scientific knowledge should be structured so as to make factual information meaningful. The way in which knowledge is structured should be influenced by the context in which the knowledge is being presented. Second, science performance depends on knowledge of facts, the ability to integrate this knowledge into larger constructs, and the capacity to use the tools, procedures, and reasoning processes of science to develop an increased understanding of the natural world. Thus, the framework called for the NAEP 1996 science assessment to include the following:

- Multiple-choice questions that assess students' knowledge of important facts and concepts and that probe their analytical reasoning skills;
- Constructed-response questions that explore students' abilities to explain, integrate, apply, reason about, plan, design, evaluate, and communicate scientific information; and
- Hands-on tasks that probe students' abilities to use materials to make observations, perform investigations, evaluate experimental results, and apply problem-solving skills.

The core of the science framework is organized along two dimensions. The first dimension divides science into three major fields: earth, physical, and life. The second dimension defines characteristic elements of knowing and doing science: conceptual understanding, scientific investigation, and practical reasoning. Each question in the assessment is categorized as measuring one of the elements of knowing and doing within one of the fields of science (e.g., scientific investigation in the context of earth science). The framework also contains two overarching domains — the nature of science and the organizing themes of science. The nature of science encompasses the historical development of science and technology, the habits of mind that characterize science, and the methods of inquiry and problem solving. It also includes the nature of technology — specifically, design issues involving the application of science to real-world problems and associated trade-offs or compromises. The themes of science include the notions of systems and their application in the scientific disciplines, models and their functioning in the development of scientific understanding, and patterns of change as they are exemplified in natural phenomena. A fuller description of the framework is provided in Appendix A.

The NAEP science assessment was developed by Educational Testing Service under a cooperative agreement with the National Center for Education Statistics. At each grade level, the NAEP 1996 science assessment consisted of 15 blocks of cognitive questions. Each student who participated in the assessment received one booklet containing three of these blocks (one of which was based on a hands-on task) and three blocks of background questions. The cognitive blocks included multiple-choice questions as well as two types of constructed-response questions: short constructed-response questions that required students to provide an answer in one or two sentences, and extended constructed-response questions that required students to provide longer answers. Answers to the constructed-response questions were evaluated using multi-level scoring guides which defined criteria for full credit, partial credit, or no credit. The background questions asked students to provide information about their background, classroom instruction, and motivation to complete the assessment. This information makes it possible to analyze and compare the performance of various subgroups of students. A more extensive discussion of the content of the assessment and of the various student, teacher, and school questionnaires is presented in Appendix A.

In 1996, a special study was conducted at grade 12 to assess students with advanced training in science. Past NAEP science assessments have been criticized for not including an adequate number of questions at advanced levels of difficulty. As a result, NAEP assessments were assumed not to have reflected what the best prepared students knew and could do in science. To be eligible for the study, students were required, at any point during the 1995-96 school year, to be enrolled in one of the following courses: (1) Advanced or Advanced Placement Biology; (2) Chemistry 2 or Advanced Placement Chemistry; or (3) Physics 1, Physics 2 without Calculus, or Advanced Placement Physics. Each student participating in the special study was administered a booklet containing three blocks of questions assessing the fields of biology, chemistry, and physics. In addition, students were also administered a block of questions from the main assessment so that the results of the special study could be linked to the results of the main assessment. The booklet also contained the same student background questions that were present in the main assessment. The results of this study will be reported at a later date.

Student Samples

The NAEP 1996 science assessment was conducted nationally at grades 4, 8, and 12, and at the state level at grade 8. For both the national and state-by-state assessments, representative samples of public and nonpublic school students were assessed. (For many of the states participating in the 1996 assessment, however, the sample of nonpublic school students was not large enough to permit the separate reporting of nonpublic school results or the combined reporting of public and nonpublic school results.) Appendix A contains information on sample sizes and participation rates for the national and state-by-state assessments.

Forty-three states, the District of Columbia, Guam, the Department of Defense Domestic Dependent Elementary and Secondary Schools (DDESS), and the overseas Department of Defense Dependents Schools (DoDDS) participated in the 1996 state-by-state assessment. (Throughout this report, participants in the state-by-state assessment are referred to as "jurisdictions.") To ensure comparability across jurisdictions, NCES established guidelines for school and student participation rates. These guidelines are included in Appendix A. Jurisdictions failing to meet any of these guidelines are so noted in the appropriate tables and figures in this report. In accordance with NCES and NAGB policies, results are not reported for jurisdictions failing to meet the initial school participation rate of 70 percent.

Figure 1.1 lists the jurisdictions that participated in the 1996 science assessment and notes those jurisdictions failing to meet one or more established participation rate guidelines for public schools. Information on public and nonpublic school participation rates is presented in Appendix A.

Figure 1.1 **Participating Jurisdictions in the NAEP 1996 State Assessment Program in Science, Grade 8**

Alabama	Indiana	Nebraska	Texas
Alaska[2]	Iowa[2]	Nevada[1]	Utah
Arizona	Kentucky	New Hampshire[1]	Vermont[2]
Arkansas[2]	Louisiana	New Jersey[1]	Virginia
California	Maine	New Mexico	Washington
Colorado	Maryland[2]	New York[2]	West Virginia
Connecticut	Massachusetts	North Carolina	Wisconsin[2]
Delaware	Michigan[2]	North Dakota	Wyoming
District of Columbia	Minnesota	Oregon	DDESS
Florida	Mississippi	Rhode Island	DoDDS
Georgia	Missouri	South Carolina[2]	Guam
Hawaii	Montana[2]	Tennessee	

[1] Failed to meet the initial school participation rate of 70 percent for public schools; public school results not reported.

[2] Failed to meet one or more participation rate guidelines for public schools; public school results reported with appropriate notation.

Reporting NAEP Results

The NAEP 1996 science assessment spans the broad field of science in each of the grades assessed. Because of the survey nature of the assessment and the breadth of the domain, each student participating cannot be expected to answer all the questions in the assessment since this would impose an unreasonable burden on students and their schools. Thus, each student was administered a portion of the assessment, and data were combined across students to report on the achievement of fourth, eighth, and twelfth graders and on the achievement of subgroups of students (e.g., subgroups defined by gender or parental education).

Student responses to the assessment questions were analyzed to determine the percentage of students responding correctly to each multiple-choice question and the percentage of students achieving each of the score categories for constructed-response questions. Item response theory (IRT) methods were used to produce scales that summarized results for each of the three fields of science (i.e., earth, physical, and life) at each grade level. These results will appear in a forthcoming NAEP report. An overall composite scale also was developed at each of grades 4, 8, and 12 by weighting the separate scales based on the relative importance of each content area in the NAEP science framework. Results presented in this report are based on this overall composite scale.

The composite scale at each grade ranges from 0 to 300. While the scale-score ranges are identical, the scale was derived independently at each grade. Also, scales were weighted differently at different grades in determining the overall scale. Therefore, average scale scores across grades cannot be compared. For example, equal scale scores on the grade 4 and grade 8 scales do not imply equal levels of science achievement.

The use of separate grade-specific reporting scales for the science assessment is consistent with the National Assessment Governing Board's 1993 policy that future NAEP assessments be developed using within-grade frameworks and that scaling be carried out within grade. The ranges of the science scales (from 0 to 300) differ by design from the 0-to-500 reporting scales used in other NAEP subject areas and were chosen to minimize confusion with other common test scales and to discourage inappropriate cross-grade comparisons. (Additional details of the scaling procedures can be found in Appendix A of this report and in the forthcoming NAEP 1996 Technical Report).

Reporting Science Achievement Level Results

A companion report, being issued by the National Assessment Governing Board, will present the NAEP 1996 science results in terms of achievement levels. As authorized by the NAEP legislation and adopted by the National Assessment Governing Board, the achievement levels are based on the Board's judgments about what are reasonable performance expectations for students in grades 4, 8, and 12 on the NAEP 1996 science assessment. The achievement levels for the NAEP 1996 science assessment were adopted on an interim basis, indicating that they may be revised when other information becomes available, such as the fourth and twelfth grade results from the Third International Mathematics and Science Study (TIMSS).

Item Maps

Another way to illustrate the range of performance on the NAEP science scale is to map questions from the assessment onto the 0-to-300 scale at each grade level. The resulting item maps are visual representations that compare questions with ability. More specifically, they indicate which questions a student at a given performance level on the NAEP scale is likely to answer correctly.[3] Figures 1.2 through 1.4 are item maps for grades 4, 8, and 12, respectively. Multiple-choice questions are shown in regular type; constructed-response questions are in italic type.[4] An example of how to interpret the item maps may be helpful. In Figure 1.2, which shows the mapping of assessment questions for grade 4, a 4-option multiple-choice question about reading the level of a liquid in a graduated cylinder maps at the 129 point on the scale. This means that fourth-grade students with science scale scores at or above 129 are likely to answer this question correctly — that is, they have at least a 74 percent chance of doing so.[5] Put slightly differently, this question is answered correctly by at least 74 of every 100 students scoring at or above the 129 scale-score level. This does not mean that students at or above the 129 scale score always answer the question correctly or that students below the 129 scale score always answer it incorrectly. Rather, the percentage of students who can successfully answer the question depends on their overall ability as measured on the NAEP science scale.

As another example, consider the constructed-response question that maps at a scale score of 194 for grade 8 (see Figure 1.3). This question concerns the differing speeds of light and sound. Scoring of responses to this question allows for partial credit by using a three-level scoring guide. Mapping a question at the 194 scale score indicates that at least 65 percent of the students performing at or above this point achieved a score of 3 ("Complete") on the question. Among students with lower scores, less than 65 percent gave complete responses to the question.

It should be noted that there were eight cross-grade blocks of questions administered in the NAEP 1996 science assessment.[6] Four of these were administered at grades 4 and 8 and four at grades 8 and 12. Thus, the question that maps at a scale-score of 164 at grade 4 (Figure 1.2) — identify organism that produces its own food — is identical to the question that maps at a scale-score of 89 at grade 8 (Figure 1.3). At fourth grade, less than 50 percent of students were likely to answer this question, whereas at eighth grade, over 90 percent were likely to do so.

---

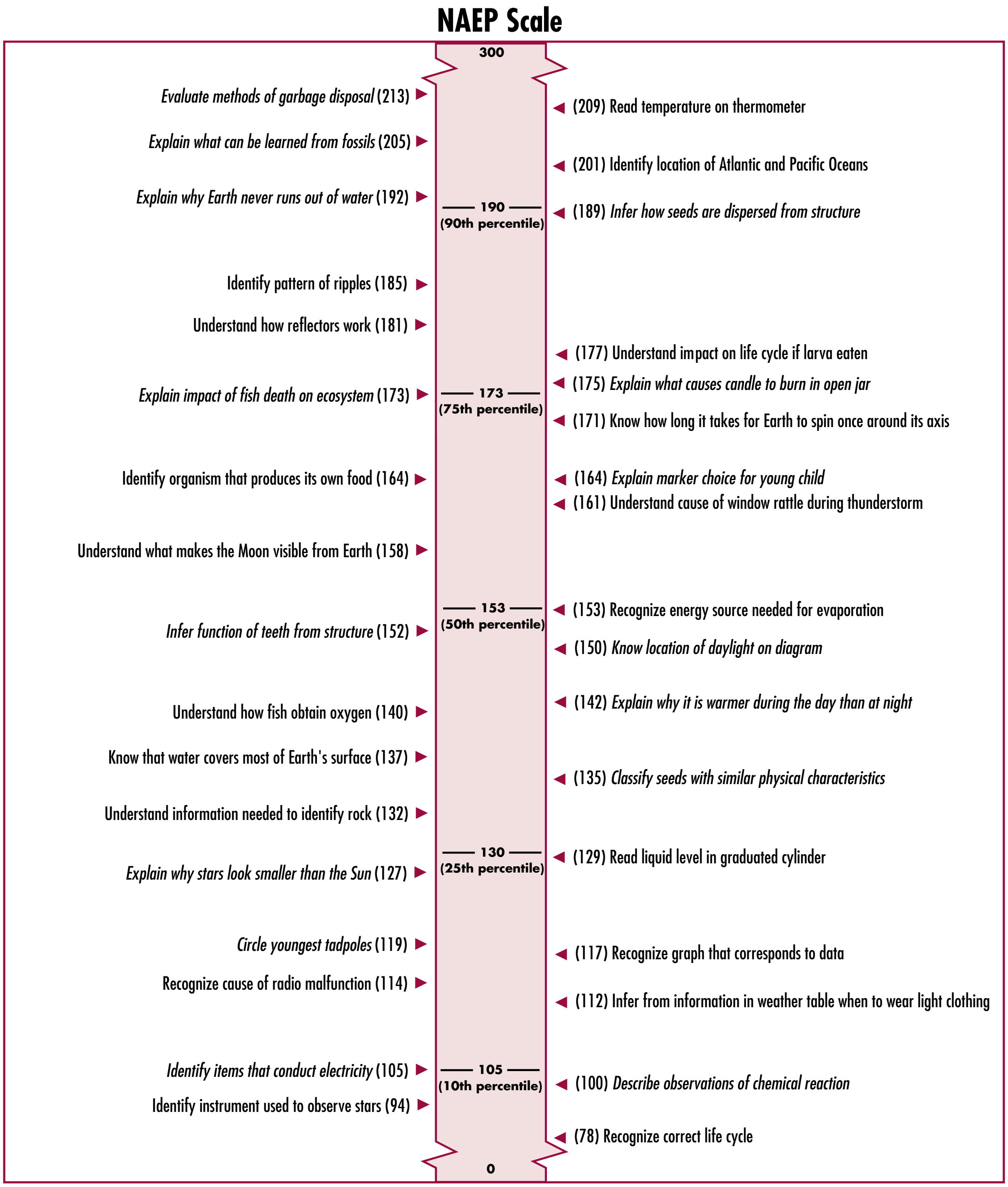
[3] Details on the procedures used to develop the item maps are provided in the forthcoming *NAEP 1996 Technical Report*. The procedures are similar to those used in past NAEP assessments.

[4] The placement of constructed-response questions is based on (1) the "mapping" of a score of 3 on a 3-point scoring guide for short constructed-response questions and (2) the "mapping" of a score of at least 3 on a 4-point scoring guide and a score of at least 4 on a 5-point scoring guide for extended constructed-response questions.

[5] For constructed-response questions, a criterion of 65 percent was used. For multiple-choice questions, the criterion was 74 percent. The use of a higher criterion for multiple-choice questions reflected students' ability to "guess" the correct answer from among the alternatives.

[6] Consistent with policy and technical considerations, the questions in cross-grade blocks were treated during scaling as distinct grade-specific sets of questions.

Figure 1.2 Map of Selected Questions on the NAEP Science Scale for Grade 4



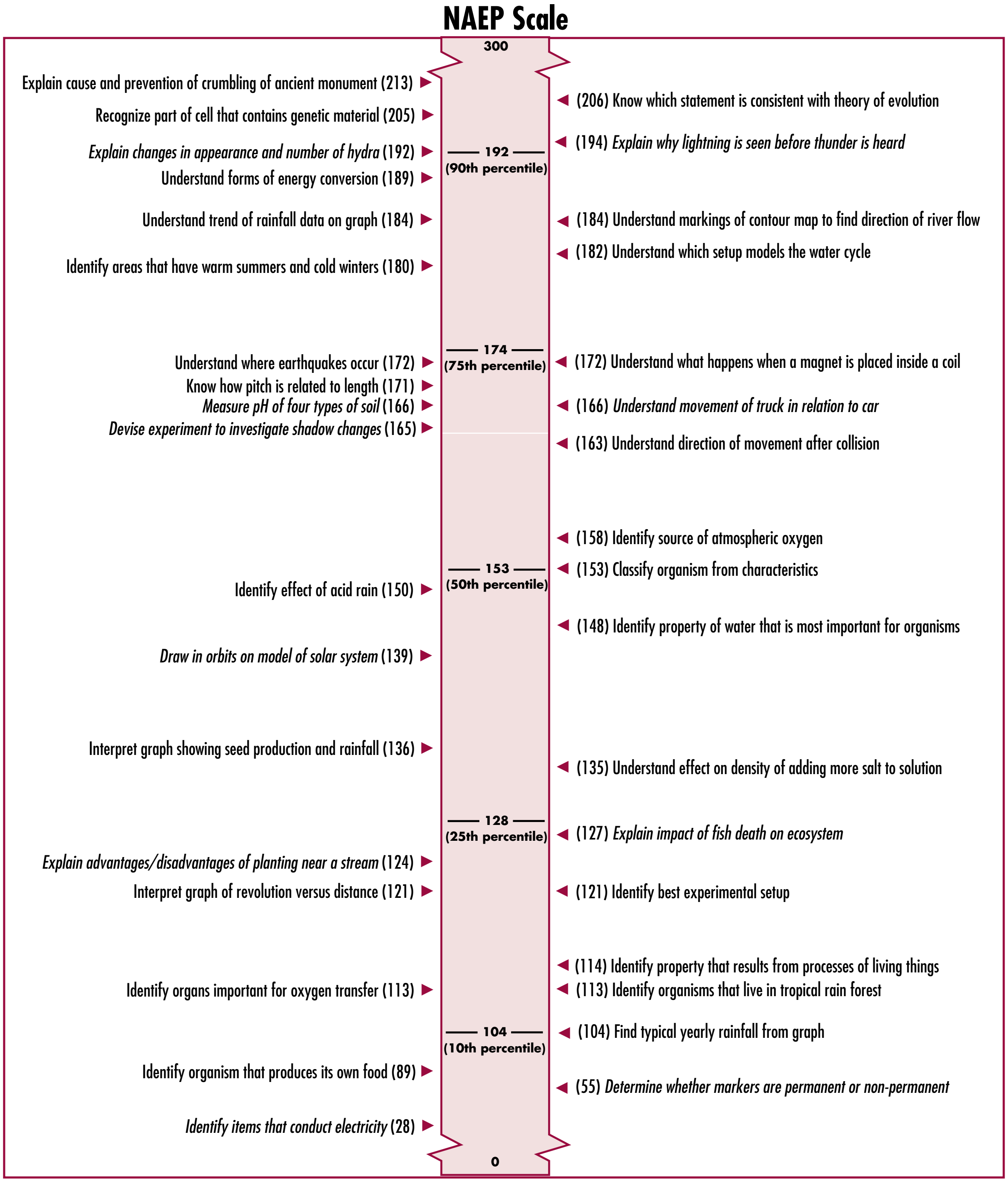
NOTE: Position of questions is approximate and an appropriate scale range is displayed for grade 4.
Italic type indicates a constructed-response question. Regular type denotes a multiple-choice question.

Each grade 4 science question was mapped onto the NAEP 0-to-300 science scale. The position of the question on the scale represents the scale score attained by students who had a 65 percent probability of reaching a given score level on a constructed-response question or a 74 percent probability of correctly answering a 4-option multiple-choice question. Only selected questions are presented. Percentiles of scale score distribution are referenced on the map.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure 1.3 Map of Selected Questions on the NAEP Science Scale for Grade 8

THE NATION'S REPORT CARD naep



NOTE: Position of questions is approximate and an appropriate scale range is displayed for grade 8.
Italic type indicates a constructed-response question. Regular type denotes a multiple-choice question.

Each grade 8 science question was mapped onto the NAEP 0-to-300 science scale. The position of the question on the scale represents the scale score attained by students who had a 65 percent probability of reaching a given score level on a constructed-response question or a 74 percent probability of correctly answering a 4-option multiple-choice question. Only selected questions are presented. Percentiles of scale score distribution are referenced on the map.

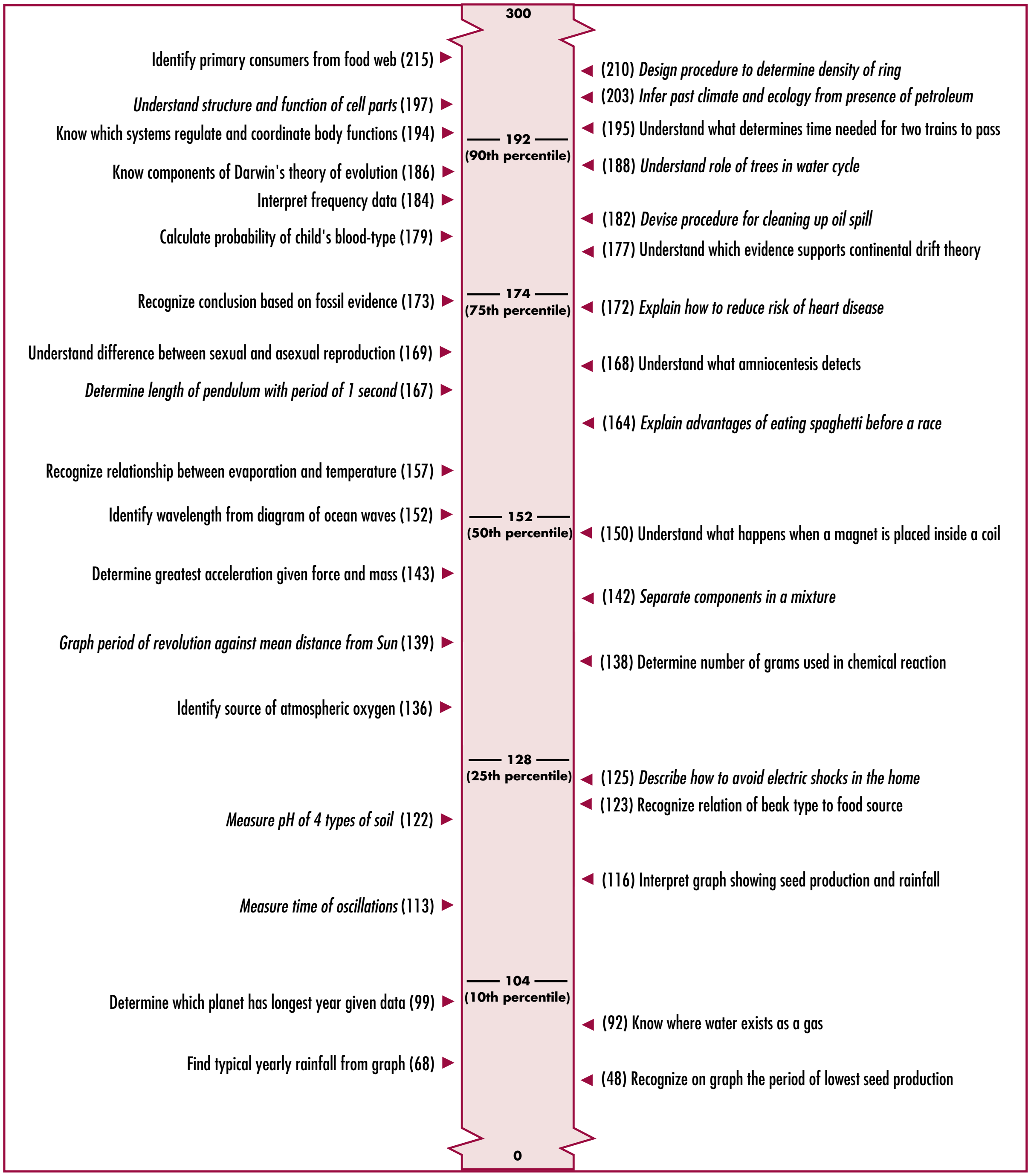
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure 1.4

Map of Selected Questions on the NAEP Science Scale for Grade 12

THE NATION'S REPORT CARD

NAEP Scale



NOTE: Position of questions is approximate and an appropriate scale range is displayed for grade 12. Italic type indicates a constructed-response question. Regular type denotes a multiple-choice question.

Each grade 12 science question was mapped onto the NAEP 0-to-300 science scale. The position of the question on the scale represents the scale score attained by students who had a 65 percent probability of reaching a given score level on a constructed-response question or a 74 percent probability of correctly answering a 4-option multiple-choice question. Only selected questions are presented. Percentiles of scale score distribution are referenced on the map.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

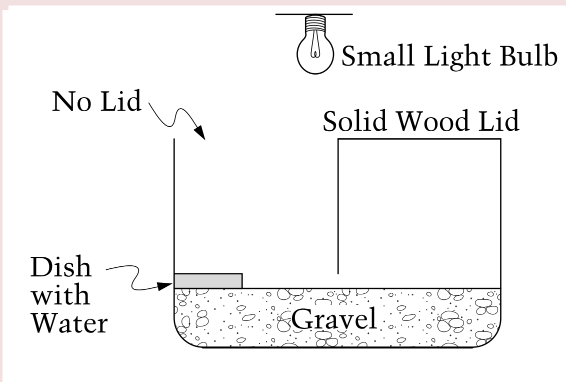
Sample Questions from the NAEP 1996 Science Assessment

As discussed earlier, the NAEP 1996 science assessment is made up of a collection of questions developed to survey the knowledge and skills specified in the assessment framework. Each student at grades 4, 8, and 12 received a mixture of multiple-choice and constructed-response questions and had the opportunity to perform a hands-on task.

Figures 1.5 through 1.7 contain samples of the types of questions used (i.e., multiple-choice, short constructed-response, and extended constructed-response). While these questions do not illustrate the breadth of the content area assessed, they do indicate the types of questions included in the assessment.

Figure 1.5 NAEP 1996 Science Sample Questions for Grade 4

Some fourth-grade students were doing a project for their science class. They were trying to find the answer to the question "Do beetles choose to live in bright light or in the shade?" The picture shows one way that a student set up an experiment to find out if beetles choose to live in bright light or in the shade.



Is this a good way to set up this experiment? Tell why or why not.

Yes cause of the wall that blocks the light. But I would put a water dish in the shad to so the water can be in wich ever side he choces.

This short constructed-response question measures Life Science and Scientific Investigation. Students' responses were scored using a three-level scoring guide that allowed for partial credit. The sample student response received the highest score, **Complete.** To receive a score of **Complete,** a student's response needed to state that the experimental design was not appropriate because there was no dish with water on the shaded side or that the experimental design was appropriate provided a dish of water was added to the shaded side.

Percentages of Fourth Graders Receiving Complete and Partial Scores	
Complete	12%
Partial	27%

Figure 1.5 (continued) NAEP 1996 Science Sample Questions for Grade 4

This multiple-choice question measures Physical Science and Conceptual Understanding and was scored as either correct or incorrect.

The correct answer is B.

Percentage of Fourth Graders Answering Correctly
85%

Kristen was listening to a portable radio one afternoon and forgot to turn it off. The next morning the radio would not work. What is the best explanation for why the radio would not work?

(A) All the radio stations stopped broadcasting.

(B) The energy stored in the batteries was all used up.

(C) It was too cold the next morning for the radio to play

(D) The radio speaker broke because it was left on for so long.

This short constructed-response question measures Earth Science and Conceptual Understanding. Students' responses were scored using a two-level scoring guide. The sample student response received a **Complete** score. To receive a score of **Complete,** a student's response needed to state that the Sun is closer than the stars and therefore looks bigger, or that stars are farther away from the Earth than the Sun and therefore look smaller.

Percentage of Fourth Graders Receiving a Complete Score
75%

Explain why many stars look smaller than the Sun even though they are really bigger than the Sun.

The sun is closer to the earth so it apears bigger then other stars father away.

Figure 1.6 NAEP 1996 Science Sample Questions for Grade 8

THE NATION'S REPORT CARD naep

A group of students took potato salad made with mayonnaise to a picnic on a very hot day. Explain how eating the potato salad could cause food poisoning.

When mayonnaise gets to hot it starts growing poisonious bactrias which can give you food poisoning

Describe something that could be done to the potato salad to prevent the people who eat it from getting food poisoning.

It can be kept in a cooler and stay cool until they want to eat it. then they should put it back in the cooler.

This short constructed-response question measures Life Science and Practical Reasoning. Students' responses were scored using a three-level scoring guide that allowed for partial credit. The sample student response received the highest score, **Complete.** To receive a score of **Complete,** a student's response needed to explain the cause of food poisoning and describe a method of preventing it.

Percentages of Eighth Graders Receiving Complete and Partial Scores	
Complete	10%
Partial	61%

Imagine that you could put popcorn kernels into an airtight popcorn popper and measure the mass of the popper with the kernels. After the popcorn has popped, the mass of the popper and the popcorn will be

- (A) less than the original mass because popped corn is less dense than the kernels are
- (B) equal to the original mass because the container is airtight
- (C) greater than the original mass because the volume of the popped corn is greater than that of the kernels.
- (D) impossible to determine accurately without weighing each piece of popcorn immediately

This multiple-choice question measures Physical Science and Conceptual Understanding and was scored as either correct or incorrect.

The correct answer is B.

Percentage of Eighth Graders Answering Correctly
26%

Figure 1.6 (continued) NAEP 1996 Science Sample Questions for Grade 8

THE NATION'S REPORT CARD naep

This extended constructed-response question measures Earth Science and Scientific Investigation. Students' responses were scored using a four-level scoring guide. The first sample student response received the highest score, **Complete.** To receive a score of **Complete,** a student's response needed to predict the relative temperature of the sand and water at noon and explain the answer. The students response also needed to give a satisfactory explanation of why the prediction might be wrong.

Percentages of Eighth Graders Receiving Complete, Essential, and Partial Scores

Complete	6%
Essential	6%
Partial	31%

The question refers to an experiment your teacher asks you to perform to compare the heating rate of soil with that of water. To do this, you are given the following materials: 2 heat lamps, 2 bins, 2 thermometers, 1 sample of soil, 1 sample of water, 1 timer.

You are instructed to heat a sample of soil and a sample of water with heat lamps, measuring the temperature of each sample once a minute for 8 minutes.

Suppose that the experiment yielded the results shown in the table below.

Time (min)	0	1	2	3	4	5	6	7	8
Soil temp (°C)	20	21	22.5	24	26	27.5	29.5	30.5	32
Water temp (°C)	20	21.5	23	23.5	24	25.5	26	27.5	28.5

At a beach that has white sand, you measure the temperature of the sand and the temperature of the seawater at 9:00 a.m. You find that both have a temperature of 16°C. If it is clear and sunny all morning, what do the data from the experiment predict about the temperature of the white sand compared to the temperature of the seawater at noon?

the sand will be hotter.

Explain your answer.

You can tell from the experiment that the soil is getting hotter faster than the water- So I believe the same thing will happen at the beach.

Explain why the prediction based on the data might be wrong.

because there are other factor to consider such as the ocean is constantly moving and thee ocean has salt in it and the ocean is heated by the sun and not a lamp.

NAEP 1996 Science Sample Questions for Grade 8

The question refers to an experiment your teacher asks you to perform to compare the heating rate of soil with that of water. To do this, you are given the following materials: 2 heat lamps, 2 bins, 2 thermometers, 1 sample of soil, 1 sample of water, 1 timer.

You are instructed to heat a sample of soil and a sample of water with heat lamps, measuring the temperature of each sample once a minute for 8 minutes.

Suppose that the experiment yielded the results shown in the table below.

Time (min)	0	1	2	3	4	5	6	7	8
Soil temp (°C)	20	21	22.5	24	26	27.5	29.5	30.5	32
Water temp (°C)	20	21.5	23	23.5	24	25.5	26	27.5	28.5

At a beach that has white sand, you measure the temperature of the sand and the temperature of the seawater at 9:00 a.m. You find that both have a temperature of 16°C. If it is clear and sunny all morning, what do the data from the experiment predict about the temperature of the white sand compared to the temperature of the seawater at noon?

It will increase more than that of the water

Explain your answer.

In the first experiment, the soil heats up faster. (Also, water has the highest Specific heat of any substance, which will take longest to heat up and cool down)

Explain why the prediction based on the data might be wrong.

I don't see why it would be wrong

The second sample student response received a score of **Essential.** To receive a score of **Essential,** a student's response needed to predict the relative temperature of the sand and water at noon and explain the answer. The response did not explain why the prediction might be wrong.

Percentages of Eighth Graders Receiving Complete, Essential, and Partial Scores

Complete	6%
Essential	6%
Partial	31%

Figure 1.7 NAEP 1996 Science Sample Questions for Grade 12

This short constructed-response question measures Earth Science and Conceptual Understanding. Students' responses were scored using a three-level scoring guide that allowed for partial credit. The sample student response received the highest score, **Complete.** To receive a score of **Complete,** a student's response needed to make correct inferences about the climate and ecology of the North Slope of Alaska millions of years ago.

Percentages of Twelfth Graders Receiving Complete and Partial Scores	
Complete	27%
Partial	21%

The petroleum fields on the North Slope of Alaska are a major energy source. What does the presence of these fields indicate about the climate and ecology of the North Slope millions of years ago?

Climate:

was warm and humed.

Ecology:

Swamp like area with much vegitation.

This multiple-choice question measures Physical Science and Conceptual Understanding and was scored as either correct or incorrect.

The correct answer is C.

Percentage of Twelfth Graders Answering Correctly
68%

Which of the following graphs shows how the rate of evaporation changes with changes in water temperature?

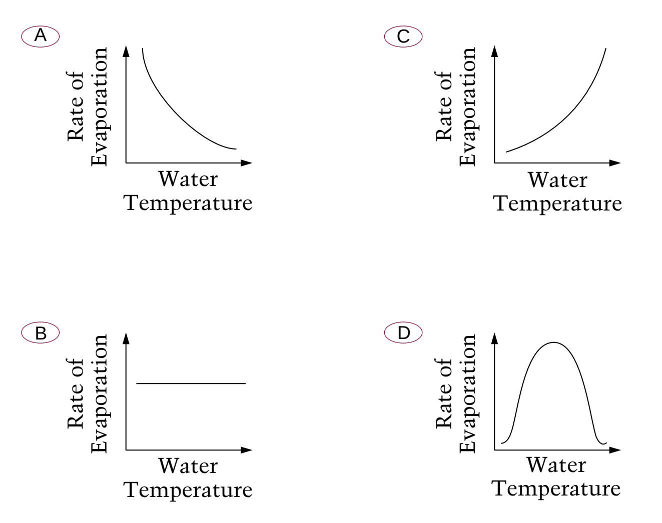


Figure 1.7 (continued) NAEP 1996 Science Sample Questions for Grade 12

A mother with attached earlobes and a father with free earlobes have 5 children — 4 boys and 1 girl. All of the children have the father's type of earlobes. What can be predicted about the genotype of the father? Construct a genetic diagram to support your prediction. What additional information, if any, would you need to determine the genotype of the father? Explain.

The genotype of the father has a dominant trait. let A be dominant trait and O be recessive trait:

	A	O
O	AO	AO
O	AO	OO

OR

	A	A
O	AO	AO
O	AO	AO

I would need the genotype of the father's parents because The father could have both dominant genes OR one dominant + the other Recessive.

This extended constructed-response question measures Life Science and Conceptual Understanding. Students' responses were scored using a four-level scoring guide. The first sample student response received the highest score, **Complete.** To receive a score of **Complete,** a student's response needed to predict the genotype of the father, provide support for the prediction, and give additional information that would be needed to determine the genotype of the father.

A mother with attached earlobes and a father with free earlobes have 5 children — 4 boys and 1 girl. All of the children have the father's type of earlobes. What can be predicted about the genotype of the father? Construct a genetic diagram to support your prediction. What additional information, if any, would you need to determine the genotype of the father? Explain.

The genotype of the father is pure. He has a pure trait for free earlobes and free earlobes are dominant over attached earlobes

Mother Father — 1

f f F F

Kids

f F

The second sample student response received a score of **Essential**. To receive a score of **Essential,** the student's response needed to predict the genotype of the father and provide support for the prediction. The response failed to state what additional information would be needed to determine the genotype of the father.

Percentages of Twelfth Graders Receiving Complete, Essential, and Partial Scores	
Complete	3%
Essential	16%
Partial	43%

Overview of the Remaining Chapters

Chapter 2 of this report presents selected results in terms of the NAEP science scales. Findings are presented for the nation, for regions, and for major reporting subgroups. In addition to the national results, state-by-state public school results are included for jurisdictions that participated in the science assessment at grade 8.

This report examines and compares the science performance of groups of students defined by demographic characteristics or by responses to background questions (e.g., males compared to females). It does not explore the relationships among combinations of these groups (e.g., White males compared to Black males). Appendix A presents detailed descriptions of the reporting subgroups.

The averages and percentages presented herein are estimates because they are based on samples rather than on all members of each population. Consequently, the results are subject to a measure of uncertainty, reflected in the standard errors of the estimates. The comparisons presented in this report are based on statistical tests that consider the magnitude of the difference between the group averages or percentages and the standard errors of those statistics. Throughout this report, differences among reporting groups are defined as significant when they are significant from a statistical perspective. The discussion of a difference as statistically significant means that observed differences in the sample are likely to reflect real differences in the population and are highly unlikely to have resulted from chance factors associated with sampling variability.[7] The term "significant," therefore, is not intended to imply a judgment about the absolute magnitude of the educational relevance of the differences. It is, rather, intended to identify statistically dependable population differences to help focus subsequent dialogue among policy makers, educators, and the public.

Chapter 3 of this report describes three hands-on assessment tasks, one each from grades 4, 8, and 12. Accompanying each description are sample questions and student responses.

The NAEP 1996 assessments in mathematics and science took several important new steps to include and accommodate more students with disabilities and limited English proficient students. Chapter 4 discusses these steps and provides an initial evaluation of their success.

Finally, this report contains appendices which support or augment the results presented. Appendix A contains an overview of the NAEP science framework and specifications, information on the national and state samples, and a detailed description of the major reporting subgroups featured in Chapter 2. The next two appendices present state-by-state results from NAEP (Appendix B) and state data from non-NAEP sources (Appendix C). Appendix D provides supporting material for Chapter 4, and Appendix E contains the standard errors for the statistics presented in this report. Detailed information about the measurement methodology and data analysis techniques will be available in the forthcoming NAEP 1996 Technical Report and the Technical Report of the NAEP 1996 State Assessment Program in Science.

---

[7] All differences reported are statistically significant at the 0.05 level with appropriate adjustments for multiple comparisons.

Cautions in Interpretations

The reader is cautioned against interpreting the relationships among subgroup averages or percentages as causal relationships. Average performance differences between two groups of students may result in part from socioeconomic and other factors. For example, differences among racial/ethnic subgroups are almost certainly associated with a broad range of socioeconomic and educational factors not discussed in this report. Similarly, differences in performance between public and nonpublic school students may be better understood by accounting for other factors such as the composition of the student body, parents' education levels, and parental involvement. Finally, student participation rates and the motivation of students, particularly twelfth graders, to perform on an assessment like NAEP should be considered when interpreting the results. (A further discussion of twelfth graders' participation rates and motivation is presented in Appendix A.)

Chapter 2

Science Scale Score Results: National and State Comparisons

The NAEP 1996 assessment gathered detailed information about the scientific knowledge and skills of our nation's fourth, eighth, and twelfth graders. In addition, state-level data were collected at grade 8 in 43 states, the District of Columbia, Guam, the Department of Defense Domestic Dependent Elementary and Secondary Schools (DDESS), and the overseas Department of Defense Dependents Schools (DoDDS).

This chapter contains detailed descriptions of the science scale-score results for the nation and for major subpopulations. The science findings show a number of consistent patterns. At all three grades there are large performance differences among racial/ethnic groups. Also, at all three grades, higher levels of parental education are generally associated with higher levels of student performance. Finally, at all three grades, groups of students who may either have low socioeconomic status or be otherwise "at risk" — specifically, those receiving Title I services or those eligible for free or reduced-price lunch — performed at substantially lower levels than other students. On the other hand, other patterns were less consistent: males outperformed females by a statistically significant amount only at grade 12.

Before proceeding with a more detailed summary of results, the reader is reminded that separate science scales were established for each of the three grades. At each grade, the average scale score for the entire population of students (e.g., all fourth graders in the nation) is 150. As a result, the scales are intended to be used to compare and describe performance *within* a grade (e.g., do eighth graders who reported that neither parent graduated from high school perform on average less well than those who reported higher levels of parental education). The scales do not allow for performance comparisons *across* grades. For example, comparisons between the average scores of fourth and eighth graders attending nonpublic schools are not meaningful.

Regional Results

For reporting purposes, the nation was divided into four regions: Northeast, Southeast, Central, and West. Each state and the District of Columbia was assigned to a region. (Appendix A provides a description of each region.) Given that each state has unique educational challenges, the regions they comprise also face varied challenges in educating their students. Across the various subjects (e.g., science, mathematics, reading, writing, history, geography, and others) assessed by NAEP, regional differences in performance typically have been found.

As presented in Table 2.1, the 1996 science assessment results also indicate regional differences in performance. Fourth-grade students attending schools in the Northeast and Central regions had higher average scores than their peers in the Southeast and West regions. At grade 8, students attending schools in the Central region had higher average scores than those in the Southeast region. Twelfth-grade students attending schools in the Central region had higher average scores than their peers in the Southeast and West. Additionally, at grade 12, students in the Northeast region outperformed students in the Southeast region. However, the fact that the average scale scores differ between regions should not lead one to believe that the performance *within* any of the regions was in any way homogeneous or monolithic. For example, at grade 12 the average score of students in the Central region is 16 points higher than the average scale score of students in the Southeast region. However, more than 25 percent of students in the Southeast region still scored well above the Central region average of 158, and more than a quarter of Central region students scored below the Southeast average of 142.

Table 2.1

Science Scale Score Results by Region: Public and Nonpublic Schools Combined

	Percentage of Students	Average Scale Score	Selected Percentiles 10th	25th	50th	75th	90th
Grade 4							
Nation	**100**	**150**	**105**	**130**	**153**	**173**	**190**
Northeast	22	156	111	137	159	179	194
Southeast	22	143	96	119	145	168	187
Central	26	156	115	137	158	177	192
West	30	146	103	126	149	169	186
Grade 8							
Nation	**100**	**150**	**104**	**128**	**153**	**174**	**192**
Northeast	22	151	105	130	153	175	194
Southeast	22	143	97	120	145	167	185
Central	24	156	111	135	159	179	197
West	32	149	102	128	153	173	191
Grade 12							
Nation	**100**	**150**	**104**	**128**	**152**	**174**	**192**
Northeast	22	154	104	130	155	179	197
Southeast	21	142	98	119	144	166	184
Central	24	158	116	138	161	181	197
West	33	147	102	126	149	170	186

NAEP science scales were developed independently for each grade assessed; therefore, results are not comparable across grades. Scale scores for all grades range from 0 to 300.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

State-Level Results

The 1996 assessment marked the first time that NAEP examined science performance at the state level. At grade 8, 43 states, the District of Columbia, Guam, the Department of Defense Domestic Dependent Elementary and Secondary Schools (DDESS), and the overseas Department of Defense Dependents Schools (DoDDS) participated in the voluntary, state-by-state NAEP science assessment. (Throughout this report, participants in the state-by-state NAEP assessment are referred to as "jurisdictions.")

In 19 of the 44 participating jurisdictions,[1] the average scale score for public school eighth graders was higher than the national average of 148, while 14 jurisdictions performed below this average. The remaining 11 jurisdictions performed at or around the national average. Comparisons of state-by-state and national results are summarized in Figure 2.1. Table 2.2 presents the overall results for the participating jurisdictions. As with the regional results, there was considerable variability within each jurisdiction as well as across jurisdictions. (Appendix B provides state-by-state results by gender, race/ethnicity, parental education, type of school, Title I participation, and eligibility for the free/reduced-price lunch program.)

Figure 2.1 **Summary of Jurisdiction Performance Relative to the Nation for Grade 8 Public Schools**

THE NATION'S REPORT CARD naep

Performed Above the National Average	Performed At or Around the National Average	Performed Below the National Average
Alaska ‡	Arizona	Alabama
Colorado	Kentucky	Arkansas ‡
Connecticut	Maryland ‡	California
DDESS	Missouri	Delaware
DoDDS	New York ‡	District of Columbia
Indiana	North Carolina	Florida
Iowa ‡	Rhode Island	Georgia
Maine	Texas	Guam
Massachusetts	Virginia	Hawaii
Michigan ‡	Washington	Louisiana
Minnesota	West Virginia	Mississippi
Montana ‡		New Mexico
Nebraska		South Carolina ‡
North Dakota		Tennessee
Oregon		
Utah		
Vermont ‡		
Wisconsin ‡		
Wyoming		

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public school participation rates (see Appendix A).

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools

DoDDS: Department of Defense Dependents Schools (Overseas)

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

[1] Several states participated but failed to meet established participation guidelines for reporting results. See Appendix A for more complete information of jurisdictions' participation rates.

Table 2.2

Science Scale Score Results by Jurisdiction for Grade 8 Public Schools

THE NATION'S REPORT CARD naep

	MEAN	10th	25th	50th	75th	90th
Performed Above the National Average						
Maine	163	128	145	164	182	196
North Dakota	162	127	146	164	181	195
Montana ‡	162	127	146	164	180	194
Wisconsin ‡	160	120	141	162	181	196
Minnesota	159	121	140	161	179	194
Iowa ‡	158	121	140	160	178	193
Wyoming	158	122	140	158	176	192
Nebraska	157	118	139	159	178	193
Vermont ‡	157	119	139	158	177	193
Massachusetts	157	114	137	160	179	196
Utah	156	120	138	158	175	190
Connecticut	155	110	135	158	179	195
DoDDS	155	118	137	157	175	190
Oregon	155	115	136	157	176	192
Colorado	155	114	136	157	176	192
Michigan ‡	153	111	133	156	176	192
Indiana	153	115	133	155	174	190
DDESS	153	117	135	153	172	188
Alaska ‡	153	111	133	156	175	192
Performed At or Around the National Average						
Missouri	151	109	132	154	172	189
Washington	150	108	130	152	172	189
Virginia	149	106	128	151	172	190
Rhode Island	149	108	129	150	171	189
Nation	**148**	**102**	**126**	**151**	**172**	**191**
Kentucky	147	107	127	149	168	185
West Virginia	147	112	129	148	166	182
North Carolina	147	104	125	148	169	187
New York ‡	146	96	122	149	172	190
Maryland ‡	145	99	123	148	170	189
Texas	145	102	123	147	169	185
Arizona	145	102	124	147	168	184
Performed Below the National Average						
Arkansas ‡	144	100	123	147	168	184
Tennessee	143	98	121	146	167	185
Florida	142	98	120	144	166	184
Georgia	142	97	120	143	166	184
Delaware	142	96	121	144	165	183
New Mexico	141	99	119	142	164	182
Alabama	139	95	117	140	163	180
South Carolina ‡	139	96	116	139	161	180
California	138	89	115	140	164	183
Hawaii	135	90	114	137	158	176
Mississippi	133	91	111	134	155	174
Louisiana	132	86	110	135	157	175
Guam	120	74	96	121	146	165
District of Columbia	113	71	90	112	135	156

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).

NAEP science scales were developed independently for each grade assessed; therefore, results are not comparable across grades. Scale scores range from 0 to 300.

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools

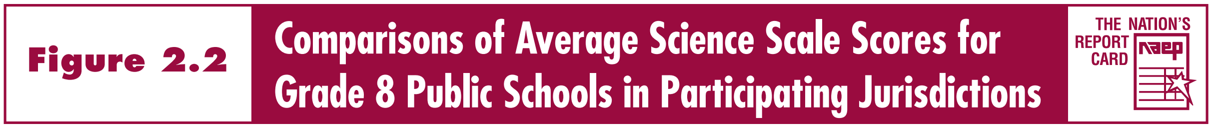
DoDDS: Department of Defense Dependents Schools (Overseas)

National results are based on the national assessment sample of public schools, not on aggregated state assessment program samples (see Appendix A).

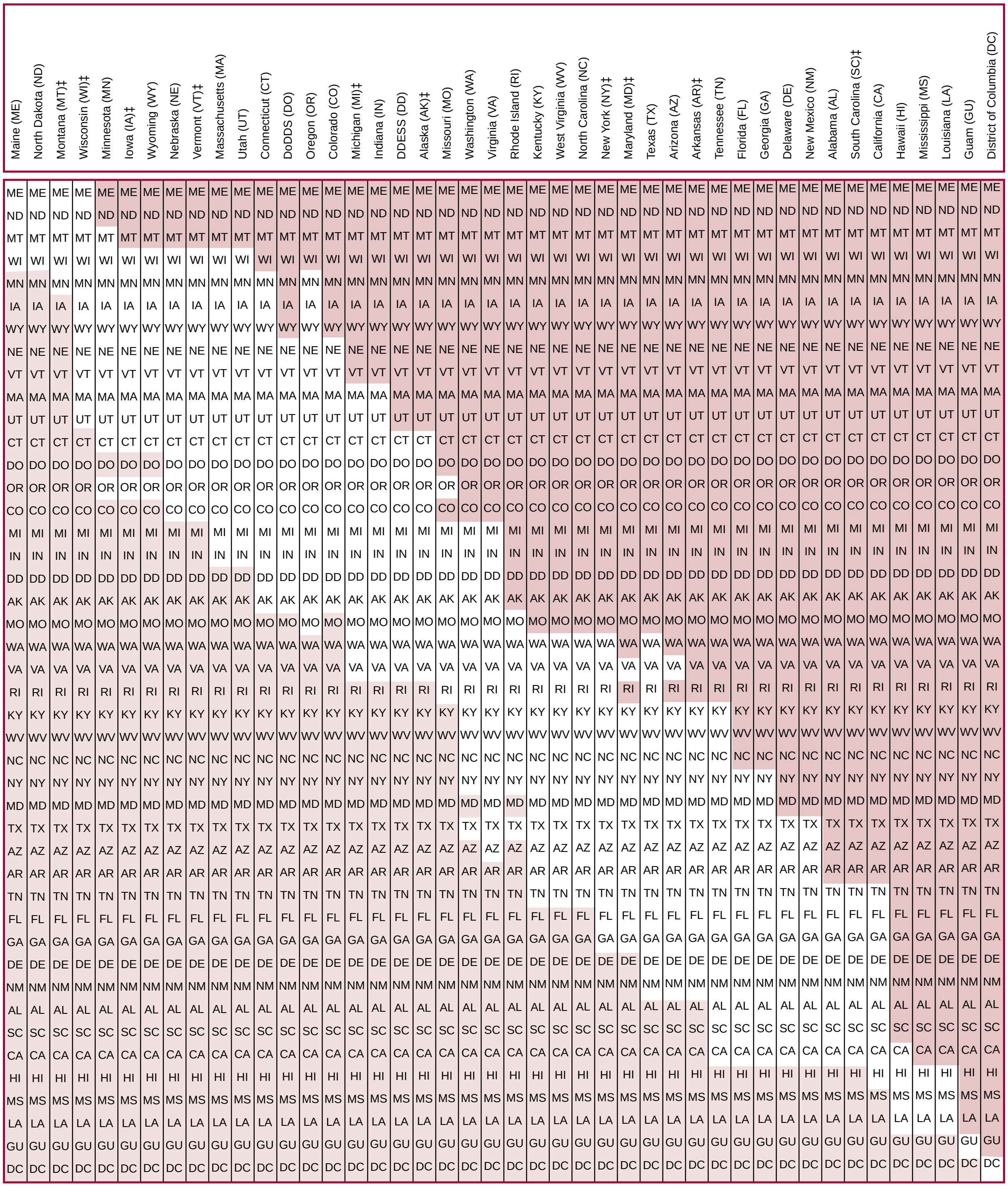
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Figure 2.2 presents another way to make valid performance comparisons across jurisdictions. This figure indicates whether or not differences between pairs of participating jurisdictions are statistically significant. For example, although average eighth-grade science scores in 1996 appear to be different between Wyoming (158) and Oregon (155), the difference is not statistically significant and may be due to chance factors such as sampling and/or measurement error.

As another example, compare the average science score for Virginia to that for each of the other 43 participating jurisdictions. Reading vertically down the column labeled "Virginia," one sees that, on average, eighth graders in Virginia scored lower than students in the jurisdictions listed from Maine through Colorado (darker shading), about the same as students in the jurisdictions listed from Michigan through Arizona (white or unshaded), and higher than students in the jurisdictions listed from Arkansas through the District of Columbia (lighter shading). From Figure 2.2, we also see that the cluster of highest performing states in 1996 consisted of four states.



Instructions: Read down the column directly under a jurisdiction name listed in the heading at the top of the chart. Match the shading intensity surrounding a jurisdiction's abbreviation to the key below to determine whether the average science scale score of this jurisdiction is higher than, the same as, or lower than the jurisdiction in the column heading.



Jurisdiction has statistically significantly higher average scale score than the jurisdiction listed at the top of the chart.

No statistically significant difference from the jurisdiction listed at the top of the chart.

Jurisdiction has statistically significantly lower average scale score than the jurisdiction listed at the top of the chart.

The between jurisdiction comparisons take into account sampling and measurement error and that each jurisdiction is being compared with every other jurisdiction. Significance is determined by an application of a multiple comparison procedure (see Appendix A).

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for school participation rates (see Appendix A).

Performance of Selected Subgroups

The following sections of this chapter present results for fourth, eighth, and twelfth graders for selected demographic subgroups. The subgroups are defined by gender, race/ethnicity, parental education, type of school, Title I participation, and eligibility for the free/reduced-price lunch program. The results from the NAEP 1996 science assessment are consistent with NAEP results in other subjects, showing considerable variability in average performance across subgroups. For example, students attending nonpublic schools outperform their peers attending public schools.

Equally important, the results also show that performance within any given subgroup varies. For example, the average science scale score of eighth graders who reported that at least one parent graduated from high school was 10 points below the national average, but more than 25 percent of these students performed better than the national average (150).

Gender

Is there a gap in science achievement between males and females? Previous NAEP science assessments have found that males outperform females in grades 8 and 12.[2] Such differences have not been observed at grade 4, however. Table 2.3 presents the NAEP 1996 science assessment results for males and females at grades 4, 8, and 12. In grades 4 and 8, the performance of males and females did not differ to a statistically significant degree. At grade 12, consistent with previous assessments, males on average outperformed their female classmates.

---

[2] Campbell, J.R., Reese, C.M., O'Sullivan, C., & Dossey, J.A., *NAEP 1994: Trends in Academic Progress* (Washington, DC: National Center for Education Statistics, 1996)

Jones, L.R., Mullis, I.V.S., Raizen, S.A., Weiss, I.R., & Weston, E.A., *The 1990 Science Report Card* (Washington, DC: Office of Educational Research and Improvement, 1992).

Table 2.3

Science Scale Score Results by Gender: Public and Nonpublic Schools Combined

THE NATION'S REPORT CARD naep

	Percentage of Students	Average Scale Score	Selected Percentiles 10th	25th	50th	75th	90th
Grade 4							
All Students	**100**	**150**	**105**	**130**	**153**	**173**	**190**
Male	50	151	105	130	154	175	191
Female	50	149	105	129	152	172	188
Grade 8							
All Students	**100**	**150**	**104**	**128**	**153**	**174**	**192**
Male	51	151	103	128	154	175	194
Female	49	149	104	128	151	172	190
Grade 12							
All Students	**100**	**150**	**104**	**128**	**152**	**174**	**192**
Male	48	152	103	129	155	178	196
Female	52	148	105	127	150	171	187

NAEP science scales were developed independently for each grade assessed; therefore, results are not comparable across grades. Scale scores for all grades range from 0 to 300.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Race/Ethnicity

Table 2.4 presents NAEP 1996 science assessment results for the following mutually exclusive categories: White, Black, Hispanic, Asian/Pacific Islander, and American Indian. NAEP creates these subgroups based on students' reports of their race/ethnicity. Past NAEP assessments have consistently reported performance differences among various racial/ethnic groups.[3] When interpreting such differences in achievement, however, confounding factors related to socioeconomics, home environment, and available educational opportunities need to be considered.[4]

Consistent with other NAEP assessments, White and Asian/Pacific Islander students have higher average science scale scores than Black and Hispanic students at all three grades (see Table 2.4). At grade 12, White students outperformed their Asian/Pacific Islander peers. At grades 4 and 8, American Indian students outperformed their Black and Hispanic peers while performing lower than their White peers. Hispanic students at grade 8 scored higher, on average, than Black students.

It was not possible to make an accurate determination of the standard error associated with the average scale score for the twelfth-grade American Indian student sample. Therefore, differences between this group and other racial/ethnic groups are not discussed.

---

[3] Campbell, J.R., Reese, C.M., O'Sullivan, C., & Dossey, J.A., *NAEP 1994: Trends in Academic Progress* (Washington, DC: National Center for Education Statistics, 1996).

[4] McKenzie, F.D., "Educational Strategies for the 1990's." *The State of Black America* (New York, NY: National Urban League, Inc., 1991).

Swinton, D.H., "The Economic Status of African Americans: Permanent Poverty and Inequity." *The State of Black America* (New York, NY: National Urban League, Inc., 1991).

Table 2.4

Science Scale Score Results by Race/Ethnicity: Public and Nonpublic Schools Combined

THE NATION'S REPORT CARD naep

	Percentage of Students	Average Scale Score	Selected Percentiles 10th	25th	50th	75th	90th
Grade 4							
All Students	**100**	**150**	**105**	**130**	**153**	**173**	**190**
Students who indicated their Race/Ethnicity as . . .							
White	69	160	123	142	161	179	194
Black	15	124	82	102	125	146	164
Hispanic	12	128	84	106	130	151	168
Asian/Pacific Islander	3	151	109	129	151	173	191
American Indian	2	144	97	120	147	171	185
Grade 8							
All Students	**100**	**150**	**104**	**128**	**153**	**174**	**192**
Students who indicated their Race/Ethnicity as . . .							
White	70	159	121	141	161	179	196
Black	14	121	82	100	121	142	159
Hispanic	12	129	84	106	130	153	171
Asian/Pacific Islander	3	152	109	130	153	175	191
American Indian	2	148	106	125	151	169	186
Grade 12							
All Students	**100**	**150**	**104**	**128**	**152**	**174**	**192**
Students who indicated their Race/Ethnicity as . . .							
White	70	159	119	139	161	180	196
Black	14	124	84	102	123	144	164
Hispanic	11	130	85	106	130	153	173
Asian/Pacific Islander	4	149	99	126	152	175	194
American Indian	1	145 !	110 !	127 !	148 !	163 !	178 !

NAEP science scales were developed independently for each grade assessed; therefore, results are not comparable across grades. Scale scores for all grades range from 0 to 300.

Percentages may not total 100 due to rounding.

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (see Appendix A).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Parents' Highest Level of Education

Each student who participated in the NAEP science assessment was asked to indicate the level of education attained by each parent. Based on this information, parents' highest level of education (as reported by students) was determined. Specifically, this reflects the highest educational level the student reported either of their parents attaining. Therefore, if a student reported that one parent graduated from college, that student's performance is included in the *Graduated from College* scale score estimates. The levels reported were: did not finish high school, graduated from high school, some education after high school, and graduated from college. An "I don't know" category was also available: 34 percent of fourth graders, 9 percent of eighth graders, and 3 percent of twelfth graders reported not knowing the educational level of either parent. Appendix A contains a discussion of the parental education variable.

In general, at all three grades, higher levels of parental education were associated with higher levels of student performance (see Table 2.5). At grades 4 and 8, students who reported that at least one parent received some education after high school or graduated from college outperformed those who reported lower levels of parental education. In addition, the average scale scores for fourth and eighth graders who reported that at least one parent graduated from high school were higher than the averages for their classmates who reported that neither parent graduated from high school. At grade 12, each increase in the level of parental education was accompanied by an increase in the average science scale score. For example, twelfth graders who reported that at least one parent graduated from college scored higher, on average, than students who reported lower levels of parental education, and students who reported that neither parent graduated from high school scored lower than all other parental education groups.

Table 2.5 Science Scale Score Results by Parents' Highest Level of Education: Public and Nonpublic Schools Combined

THE NATION'S REPORT CARD naep

	Percentage of Students	Average Scale Score	Selected Percentiles 10th	25th	50th	75th	90th
Grade 4							
All Students	**100**	**150**	**105**	**130**	**153**	**173**	**190**
Students who reported their parents' highest level of education as . . .							
Did Not Finish High School	4	136	92	115	139	158	173
Graduated From High School	14	146	101	126	149	168	185
Some Education After High School	8	155	112	138	159	178	193
Graduated From College	41	158	112	137	161	181	196
I Don't Know	34	144	101	124	146	165	183
Grade 8							
All Students	**100**	**150**	**104**	**128**	**153**	**174**	**192**
Students who reported their parents' highest level of education as . . .							
Did Not Finish High School	6	131	87	109	134	153	170
Graduated From High School	20	140	98	119	142	163	181
Some Education After High School	20	155	114	137	158	176	191
Graduated From College	45	159	115	139	161	181	199
I Don't Know	9	134	89	110	136	158	174
Grade 12							
All Students	**100**	**150**	**104**	**128**	**152**	**174**	**192**
Students who reported their parents' highest level of education as . . .							
Did Not Finish High School	7	123	82	102	124	146	162
Graduated From High School	18	140	98	119	141	162	181
Some Education After High School	26	151	109	131	153	173	189
Graduated From College	47	160	117	140	163	182	198
I Don't Know	3	116	75	94	113	136	163

NAEP science scales were developed independently for each grade assessed; therefore, results are not comparable across grades. Scale scores for all grades range from 0 to 300.

Percentages may not total 100 due to rounding.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Type of School

Approximately 90 percent of the nation's fourth-, eighth-, and twelfth-grade students attend public schools. The remainder attend Catholic and other private schools (i.e., nonpublic schools). Past NAEP results across a variety of subjects have consistently shown students attending nonpublic schools outperforming those attending public schools. As presented in Table 2.6, the NAEP 1996 science results again repeat this pattern at all three grade levels. However, the difference in average scale scores between students attending public and nonpublic schools is far less than the range of scores within each group. At all three grades, a substantial percentage of students (i.e., more than 25 percent) attending public schools score above the average for students attending nonpublic schools. Conversely, a comparable percentage of students attending nonpublic schools score below the average for students attending public schools.

Caution should be taken not to use these data to make simplistic inferences about the relative effectiveness of public and nonpublic schools. Average performance differences between the two types of schools may be partly related to socioeconomic or sociological factors, such as parental education or parental involvement. To get a clearer picture of the differences between public and nonpublic schools, more in-depth investigations are needed.

Table 2.6 Science Scale Score Results by Type of School

	Percentage of Students	Average Scale Score	Selected Percentiles 10th	25th	50th	75th	90th
Grade 4							
All Students	**100**	**150**	**105**	**130**	**153**	**173**	**190**
Students who attend . . .							
Public Schools	88	148	103	127	151	172	188
Nonpublic Schools:	12	163	126	145	165	182	197
Catholic Schools	8	163	127	146	165	182	197
Other Private Schools	5	163 !	124 !	145 !	164 !	183 !	197 !
Grade 8							
All Students	**100**	**150**	**104**	**128**	**153**	**174**	**192**
Students who attend . . .							
Public Schools	89	148	102	126	151	172	191
Nonpublic Schools:	11	162	123	143	164	182	199
Catholic Schools	7	163	125	144	165	182	199
Other Private Schools	5	161 !	118 !	142 !	164 !	183 !	199 !
Grade 12							
All Students	**100**	**150**	**104**	**128**	**152**	**174**	**192**
Students who attend . . .							
Public Schools	88	149	103	126	151	174	192
Nonpublic Schools:	12	155	115	136	156	175	191
Catholic Schools	8	154	116	136	155	174	190
Other Private Schools	4	155 !	111 !	137 !	158 !	177 !	193 !

NAEP science scales were developed independently for each grade assessed; therefore, results are not comparable across grades. Scale scores for all grades range from 0 to 300.

Percentages of students attending Catholic Schools and Other Private Schools may not total the percentage for Nonpublic Schools due to rounding.

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (see Appendix A).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Participation in Title I Programs

The Improving America's Schools Act of 1994 (P.L. 103-382) reauthorized the Elementary and Secondary Education Act of 1965 (ESEA). Title I, Part A of ESEA provides local education agencies with financial assistance to meet the educational needs of children who are performing below grade level and who are economically disadvantaged.[5] Title I programs are designed to help disadvantaged students meet challenging academic performance standards. Through Title I, schools are assisted in improving teaching and learning and in providing students with opportunities to acquire the knowledge and skills outlined in their state's curriculum content and performance standards. In schools in high poverty areas all children in the school may benefit from participation in schoolwide Title I programs. Title I funding supports state and local education reform efforts and promotes the coordination of resources to improve education for all students.

Nationally, 22 percent of fourth graders and 12 percent of eighth graders were receiving Title I services during the 1995-96 academic year. As would be expected, these students scored lower in the assessment, on average, than did other students (see Table 2.7). At grade 12, the vast majority of students were not participating in Title I programs. The differences between twelfth graders who were and were not participating are not discussed here because the nature of the sample does not allow for accurate estimation of the variability of the average scale score for Title I participants.

The Title I information collected by NAEP refers to current participation in Title I programs. Thus, students who participated in such services in the past but who do not currently receive services are not identified as Title I participants. Differences in assessment results between students who receive Title I services and those who do not should not be used as an evaluation of Title I programs. Typically, Title I services are intended for students who score poorly on assessments. To properly evaluate Title I programs, the performance of students participating in such programs must be monitored and assessed over time.[6]

[5] U.S. Department of Education, Office of Elementary and Secondary Compensatory Education Programs, *Improving Basic Programs Operated by Local Education Agencies* (Washington, DC: U.S. Department of Education, 1996).

[6] For a study of mathematics performance of Title I students in 1991-92, see U.S. Department of Education, *PROSPECTS: The Congressionally Mandated Study of Educational Growth and Opportunity, Interim Report: Language Minority and Limited English Proficient Students* (Washington, DC: U.S. Department of Education, 1995).

Table 2.7 Science Scale Score Results by Participation in Title I Programs: Public and Nonpublic Schools Combined

	Percentage of Students	Average Scale Score	Selected Percentiles 10th	25th	50th	75th	90th
Grade 4							
All Students	**100**	**150**	**105**	**130**	**153**	**173**	**190**
Participated	22	126	84	104	127	148	165
Did not participate	78	157	117	139	159	178	193
Grade 8							
All Students	**100**	**150**	**104**	**128**	**153**	**174**	**192**
Participated	12	126	82	102	126	151	169
Did not participate	88	153	109	133	156	176	194
Grade 12							
All Students	**100**	**150**	**104**	**128**	**152**	**174**	**192**
Participated	3	118 !	79 !	96 !	118 !	139 !	158 !
Did not participate	97	151	106	129	153	175	192

NAEP science scales were developed independently for each grade assessed; therefore, results are not comparable across grades. Scale scores for all grades range from 0 to 300.

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (see Appendix A).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Eligibility for the Free/Reduced-Price Lunch Program

The free/reduced-price lunch component of the National School Lunch Program (NSLP), offered through the U.S. Department of Agriculture (USDA), is designed to ensure that children near or below the poverty level receive nourishing meals.[7] This program is available to students attending public schools, nonprofit private schools, and residential child care institutions. Eligibility for free or reduced-priced meals is determined through the USDA's Income Eligibility Guidelines. Eligibility for the free/reduced-priced lunch program is included in NAEP as an indicator of poverty.

As shown in Table 2.8, the average science scale scores for the nation's poorest fourth-, eighth-, and twelfth-grade students (i.e., those who were eligible for the free/reduced-price lunch program) were lower than those of students who were not eligible. It should be noted that for 12 percent of fourth graders, 23 percent of eighth graders, and 21 percent of twelfth graders, information on their eligibility for this program was not available.

Table 2.8

Science Scale Score Results by Eligibility for the Free/Reduced-Price Lunch Program: Public and Nonpublic Schools Combined

THE NATION'S REPORT CARD naep

	Percentage of Students	Average Scale Score	Selected Percentiles 10th	25th	50th	75th	90th
Grade 4							
All Students	**100**	**150**	**105**	**130**	**153**	**173**	**190**
Eligible	35	133	88	110	135	157	175
Not Eligible	53	159	122	141	160	178	194
Information Not Available	12	161	117	141	163	183	200
Grade 8							
All Students	**100**	**150**	**104**	**128**	**153**	**174**	**192**
Eligible	26	133	88	108	133	157	176
Not Eligible	51	156	116	137	158	177	194
Information Not Available	23	156	112	136	159	180	198
Grade 12							
All Students	**100**	**150**	**104**	**128**	**152**	**174**	**192**
Eligible	11	125	81	101	124	150	171
Not Eligible	68	154	111	134	156	177	193
Information Not Available	21	150	104	127	151	175	192

NAEP science scales were developed independently for each grade assessed; therefore, results are not comparable across grades. Scale scores for all grades range from 0 to 300.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

[7] U.S. General Services Administration, *Catalog of Federal Domestic Assistance* (Washington, DC: Executive Office of the President, Office of Management and Budget, 1995).

Summary

The preceding sections provided a detailed picture of science achievement for students in various subgroups defined by region, gender, race/ethnicity, parental education, type of school, Title I participation, and eligibility for the free/reduced-price lunch program. Although results differed slightly by grade the following patterns emerged.

- At each grade, students attending schools in the Central region had higher average scores than their peers in the Southeast region.
- In grades 4 and 8, the performance of male and female students did not vary by a statistically significant amount. However, at grade 12, male students outperformed their female classmates.
- Consistent with past NAEP assessments, White and Asian/Pacific Islander students had higher average science scale scores than Black and Hispanic students at all three grades.
- In general, at all three grades higher levels of parental education were associated with higher levels of student performance.
- Fourth, eighth, and twelfth graders attending nonpublic schools outperformed their peers attending public schools.
- At grades 4 and 8, students currently participating in Title I programs scored lower, on average, than those not participating in such programs. (A comparison at grade 12 was not possible due to sample limitations.)
- Across the three grades, students eligible for the free/reduced-price lunch program scored lower, on average, than their classmates who were not eligible.

Chapter 3

Student Performance on Hands-On Science Tasks

Introduction

A number of goals for science education have been put forward in a series of reports authored by government agencies and professional societies over the last 15 years.[1] These goals include acquisition of a core of scientific understanding, ability to apply science knowledge in practical ways, familiarity with experimental design, and the ability to carry out scientific experiments. The reports also offered recommendations for the science curricula and instruction needed to achieve these goals. One recommendation was to encourage active student participation in hands-on science, learning in cooperative groups, and completing sustained projects.[2]

A 1993 national survey indicated that science teachers devote 21 to 26 percent of class time to hands-on or manipulative activities.[3] While research on the relationship between exposure to hands-on science tasks and overall science performance is sparse and inconclusive, a recent study has demonstrated a positive relationship for eighth-grade students between the frequency of hands-on activities and their performance on a standardized assessment.[4]

---

[1] National Science Board Commission on Precollege Education in Mathematics, Science, and Technology, *Educating America for the 21st Century* (Washington, DC: National Science Foundation, 1983).

American Association for the Advancement of Science, *Science For All Americans: A Project 2061 Report On Literacy Goals in Science, Mathematics, and Technology* (Washington, DC: American Association for the Advancement of Science, 1989).

Aldridge, B.G., *Essential Changes in Secondary School Science: Scope, Sequence, and Coordination* (Washington DC: National Science Teachers Association, 1989).

National Research Council, *Fulfilling the Promise: Biology Education in the Nation's Schools* (Washington, DC: National Academy Press, 1990).

[2] *Science Framework for the 1996 National Assessment of Educational Progress* (Washington, DC: National Assessment Governing Board, 1995).

[3] Blank, R.K., & Gruebel, D., *State Indicators of Science and Mathematics Education* (Washington, DC: Council of Chief State School Officers, 1995)

[4] Stohr-Hunt, P.M., "An Analysis of Frequency of Hands-On Experience and Science Achievement." *Journal of Research in Science Teaching* 33, 101-109 (1996).

NAEP included assessments of higher-order thinking skills in science and mathematics as early as 1986 through a pilot assessment that required students to work on various hands-on tasks. Although the NAEP 1990 science assessment measured skills that were integral to scientific investigation,[5] hands-on tasks were not included. When the 1996 science framework[6] was developed in the early 1990s, it took into account the current reforms in science education by specifying three question types that probed understanding of conceptual and reasoning skills: performance exercises, constructed-response questions, and multiple-choice questions. It was envisaged that in the performance exercises, students would manipulate selected physical objects and try to solve a scientific problem using the objects before them. Hands-on tasks that met these criteria were developed for the 1996 science assessment, and each student who participated in the assessment was given an opportunity to conduct one of them.

NAEP Hands-On Science Tasks

Ten hands-on tasks were administered in the NAEP 1996 science assessment. Each task was designed to use materials to perform an investigation, make observations, evaluate experimental results, and apply problem-solving skills. In addition, tasks shared the following characteristics.

- Diagrams were included to guide students through the procedures.
- Multiple-choice and constructed-response questions were embedded throughout the tasks.
- Scientific investigation was integrated with conceptual understanding and practical reasoning.

The creation of the hands-on tasks presented special challenges. Since the assessment was administered in a variety of settings, ranging from laboratories to cafeterias, all of the required equipment necessary to conduct each task had to be provided in a self-contained kit produced according to standard specifications to ensure uniformity. There were some limitations on materials and equipment. For example, live materials (with the exception of seeds) and equipment that required an electric outlet were not used. Safety was also an important concern and was addressed in a number of ways. States' safety regulations were considered; no toxic or corrosive chemicals were used; assessment administrators were trained in appropriate laboratory safety; and students were provided with goggles for some of the tasks.

Three hands-on tasks are described in this chapter. Samples of assessment questions and student responses are shown together with scoring criteria and question-level results.

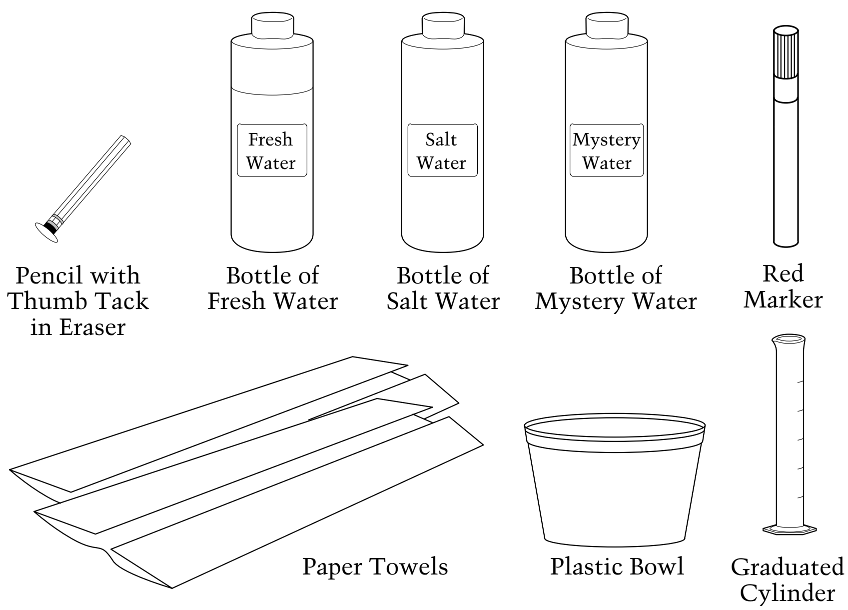
[5] *Science Objectives: 1990 Assessment* (Princeton, NJ: The National Assessment of Educational Progress, 1989).

[6] *Science Framework for the 1996 National Assessment of Educational Progress* (Washington, DC: The National Assessment Governing Board, 1995).

FLOATING PENCIL

Using a Pencil to Test Fresh and Salt Water

You have been given a bag with some things in it that you will work with during the next 20 minutes. Take all of the things out of the bag and put them on your desk. Now look at the picture below. Do you have everything that is shown in the picture? If you are missing anything, raise your hand and you will be given the things you need.



An instrument constructed from a pencil and thumbtack served as a hydrometer in this task. Students were asked to observe, measure, and compare the lengths of a portion of pencil, marked with calibrations for ease of measurement, that floated above the water surface in fresh water and salt water. The students then determined if an unknown water sample was fresh water or salt water and predicted how the addition of more salt to the salt water would affect the floating pencil. The task assessed students' ability to make simple observations, measure volume using a graduated cylinder, measure length using a ruler, apply observations and measurements to test an unknown, make generalized inferences from observations, and apply understanding to an everyday situation.

Figure 3.1 presents a short constructed-response question that asks students to use the floating-pencil test to find out if the water in a bottle labeled "Mystery Water" is fresh water or salt water and explain how they are able to tell. This question was presented towards the end of the task after students had measured the height of the pencil above the fresh water, salt water, and the mystery water. Responses to this question were scored according to a three-level guide: *Complete, Partial, or Incorrect.* Figure 3.1 also presents a sample of a student response that received a score of *Complete.* The response received a score of *Complete* because the mystery water was identified and the explanation specifically referred to the level the fresh water and the mystery water reached on the calibrated pencil. Twenty-eight percent of students were able to correctly identify the mystery water and give a satisfactory explanation.

Figure 3.1 — **Sample Question One from the Grade 4 Hands-On Task: Floating Pencil**

THE NATION'S REPORT CARD naep

Students' responses were scored using a three-level scoring guide that allowed for partial credit. The sample student response received the highest score, **Complete** because it stated that the mystery water was fresh water and gave a satisfactory explanation that referred to observations made when doing the hands-on task.

Is the mystery water fresh water or is it salt water?

fresh water

How can you tell what the mystery water is?

because in the fresh water it went to A and it went to A again

Percentages of Fourth Graders Receiving Complete and Partial Scores	
Complete	28%
Partial	45%

Figure 3.2 presents a short constructed-response question that asks students to apply their observations of the behavior of a pencil in different solutions to a real-world situation (swimming in salt water and fresh water). This question was presented at the end of the task after students had measured the height of the pencil above the fresh water, salt water, and the mystery water and determined what the mystery water was. Responses to this question were scored according to a three-level guide: *Complete, Partial,* or *Incorrect.* Figure 3.2 also presents a sample of a student response that received a score of *Complete.* The ocean was correctly identified and the explanation referred to information learned by performing the hands-on-task. Fourteen percent of students were able to apply their findings.

Figure 3.2

Sample Question Two from the Grade 4 Hands-On Task: Floating Pencil

THE NATION'S REPORT CARD

Students' responses were scored using a three-level scoring guide that allowed for partial credit. The sample student response received the highest score, **Complete** because it stated that it was easier to stay afloat in the ocean and gave a satisfactory explanation that referred to information learned while conducting the hands-on task.

When people are swimming, is it easier for them to stay afloat in the ocean or in a freshwater lake?

ocean

Explain your answer.

The ocean has salt water and the pencoil stayed higher a'float in salt water than fresh water.

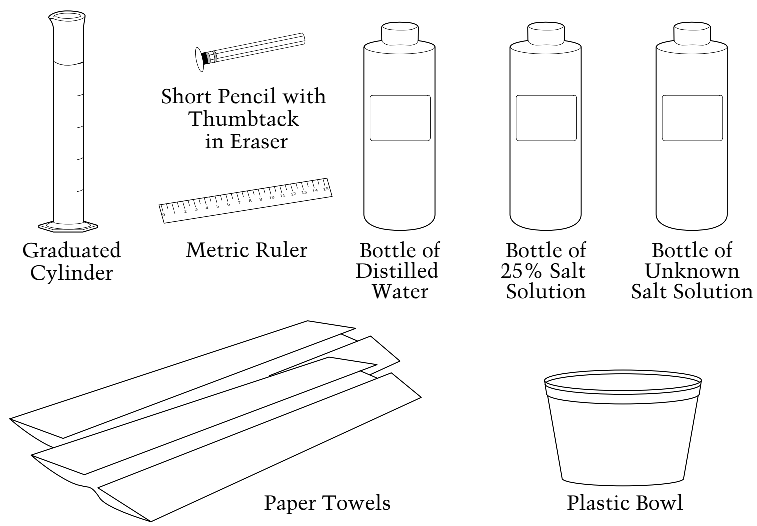
Percentages of Fourth Graders Receiving Complete and Partial Scores	
Complete	14%
Partial	29%

Grade 8: Task Summaries and Sample Questions

SALT SOLUTIONS

Estimating the Salt Concentration of an Unknown Salt Solution Using the "Floating Pencil Test"

For this task, you have been given a kit that contains materials that you will use to perform an investigation during the next 30 minutes. Please open your kit now and use the following diagram to check that all of the materials in the diagram are included in your kit. If any materials are missing, raise your hand and the administrator will provide you with the materials that you need.



An instrument constructed from a pencil and thumbtack served as a hydrometer in this task. Students were asked to observe, measure, and compare the lengths of a portion of pencil, marked with calibrations for ease of measurement, that floated above the surface in distilled water and in a 25% salt solution. Based on these observations, the students were asked to predict how the addition of more salt to the salt solution would affect the floating pencil. Students then measured the length of the pencil that floated above the surface of a solution of unknown salt concentration and used the results of their previous observations to estimate the salt concentration of the unknown solution. The task assessed students' ability to make simple observations, measure length using a ruler, apply observations to an unknown, draw a graph, interpolate from graphical data, and make a generalized inference from observations. The task also assessed students' understanding of the value of performing multiple trials of the same procedure.

Figure 3.3 shows four constructed-response questions administered at grade 8. Students were asked to measure the height of a pencil that floated above the surface of three solutions: distilled water, a 25% salt solution, and a solution containing an unknown concentration of salt. They recorded two measurements for each solution and calculated the average of each pair of readings. Students were then asked to graph their results and use the graph to interpolate the concentration of salt in the unknown solution. The data recorded in the table were scored using two guides. The ability to accurately observe, measure, and record the length of the pencil floating above each solution was scored according to a four-level guide: *Complete, Essential, Partial,* or *Incorrect*. A score of *Complete* required that all three sets of measurements agreed within a tolerance ± 0.2 cm or ± 1/16 inch and that the relative order of measurements was correct. The ability to accurately calculate the average of two measurements was scored according to a three-level guide: *Complete, Partial,* or *Incorrect*. A score of *Complete* required that all three of the student-calculated averages were within 0.1 cm or 1/32 inch of the correct averages as calculated from the student's results.

The question that required students to draw a graph was scored according to a three-level guide: *Complete, Partial,* or *Incorrect*. A graph that was correctly plotted from student data received a score of *Complete*. Finally, students had to use the graph to interpolate the salt concentration of the unknown solution. This was scored according to a four-level guide: *Complete, Essential, Partial*, or *Incorrect*. In order to receive a score of *Complete* the student's response had to show a salt concentration consistent with student data and give a satisfactory explanation. It should be noted that if the averages were calculated incorrectly, but the graph based on these incorrectly derived averages was drawn correctly, the graph would receive a score of *Complete*. Similarly, if the graph was incorrectly drawn, but the interpolation was done correctly with a satisfactory explanation, this response would also receive a score of *Complete*.

Figure 3.3 includes sample responses from two students as well as the percentages of students receiving scores of *Complete, Essential,* and *Partial* and scores of *Complete* and *Partial.*

Figure 3.3 Sample Questions from the Grade 8 Hands-On Task: Salt Solution

Measurement: Students' responses were scored using a four-level scoring guide. The sample student response received the highest score, **Complete** because the three sets of measurements agreed within tolerance and were in the correct relative order.

Percentages of Eighth Graders Receiving the Following Scores		
Complete	**Essential**	**Partial**
42%	**16%**	**21%**

Average: Students' responses were scored using a three-level scoring guide. The sample student response received a score of **Complete** because the three averages were correctly calculated.

Percentages of Eighth Graders Receiving the Following Scores	
Complete	**Partial**
57%	**22%**

Graph: Students' responses were scored using a three-level scoring guide. The sample student response received a score of **Complete** because the two data points were plotted correctly.

Percentages of Eighth Graders Receiving the Following Scores	
Complete	**Partial**
28%	**19%**

Interpolation: Students' responses were scored using a four-level scoring guide. The sample student response received a score of **Complete** because the concentration of salt in the unknown solution was interpolated correctly and a satisfactory explanation was given.

Percentages of Eighth Graders Receiving the Following Scores		
Complete	**Essential**	**Partial**
20%	**8%**	**16%**

Now take the pencil out of the water and dry it with a paper towel. Use the ruler to measure the length of the pencil that was above the water. Record the length in Table 1 below under **Measurement 1.**

TABLE 1

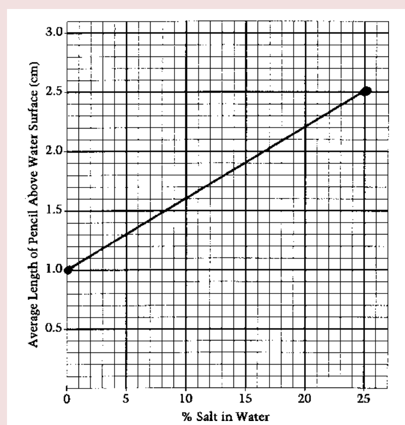
Type of Solution	Length of Pencil Above Water Surface (cm)		
	Measurement 1	Measurement 2	Average
Distilled Water	1 cm	1cm	1 cm
Salt Solution	2½cm	2½cm	2½c m
Unknown Salt Solution	2 cm	2 cm	2 cm

Now place the pencil back in the distilled water and repeat steps 2 and 3. Record your measurement in Table 1 under **Measurement 2**.

Calculate the average of Measurements 1 and 2 and record the result in the data table.

(You can calculate the average by adding Measurement 1 + Measurement 2 and then dividing by two.)

On the graph below, plot the average values you obtained for the distilled water and the 25% salt solution. Draw a straight line between the two data points. Assume that this line represents the relationship between the length of pencil that is above the water surface and the concentration of salt in the water.



Based on the graph that you plotted, what is the salt concentration of the unknown solution? about 16%

Explain how you determined your answer.

with the line connecting the distilled water and the salt solution I measured 2 cm. The measurement connects to the line at about 16% so the unknown solution is about 16%

Sample Questions from the Grade 8 Hands-On Task: Salt Solution

Now take the pencil out of the water and dry it with a paper towel. Use the ruler to measure the length of the pencil that was above the water. Record the length in Table 1 below under **Measurement 1**.

TABLE 1

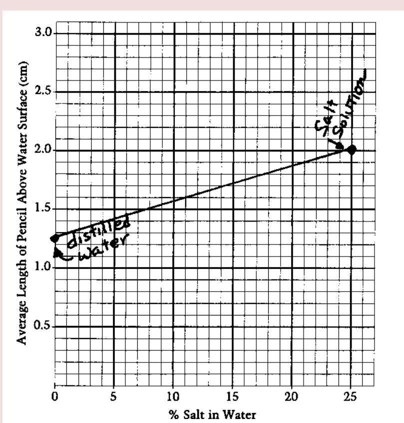
Type of Solution	Length of Pencil Above Water Surface (cm)		
	Measurement 1	Measurement 2	Average
Distilled Water	1 metric cm	1/2 metric cm	1 1/2 metric cm
Salt Solution	2 metric cm	2 metric cm	2 metric cm
Unknown Salt Solution	1 1/2 metric cm	1 1/2 metric cm	2 2/2 metric cm

Now place the pencil back in the distilled water and repeat steps 2 and 3. Record your measurement in Table 1 under **Measurement 2**.

Calculate the average of Measurements 1 and 2 and record the result in the data table.

(You can calculate the average by adding Measurement 1 + Measurement 2 and then dividing by two.)

On the graph below, plot the average values you obtained for the distilled water and the 25% salt solution. Draw a straight line between the two data points. Assume that this line represents the relationship between the length of pencil that is above the water surface and the concentration of salt in the water.



Based on the graph that you plotted, what is the salt concentration of the unknown solution? About 12%

Explain how you determined your answer.

Becase I split it in half between the two

Measurement: Students' responses were scored using a four-level scoring guide. The sample student response received a score of **Partial** because only two sets of measurements agreed within tolerance.

Percentages of Eighth Graders Receiving the Following Scores		
Complete	Essential	Partial
42%	16%	21%

Average: Students' responses were scored using a three-level scoring guide. The sample student response received a score of **Partial** because only one average was correctly calculated.

Percentages of Eighth Graders Receiving the Following Scores	
Complete	Partial
57%	22%

Graph: Students' responses were scored according to a three-level scoring guide. The sample student response received a score of **Partial** because only one of the two data points was plotted correctly.

Percentages of Eighth Graders Receiving the Following Scores	
Complete	Partial
28%	19%

Interpolation: Students' responses were scored according to a four-level scoring guide. The sample student response received a score of **Incorrect** because the student did not give a response consistent with the data and the explanation was unsatisfactory.

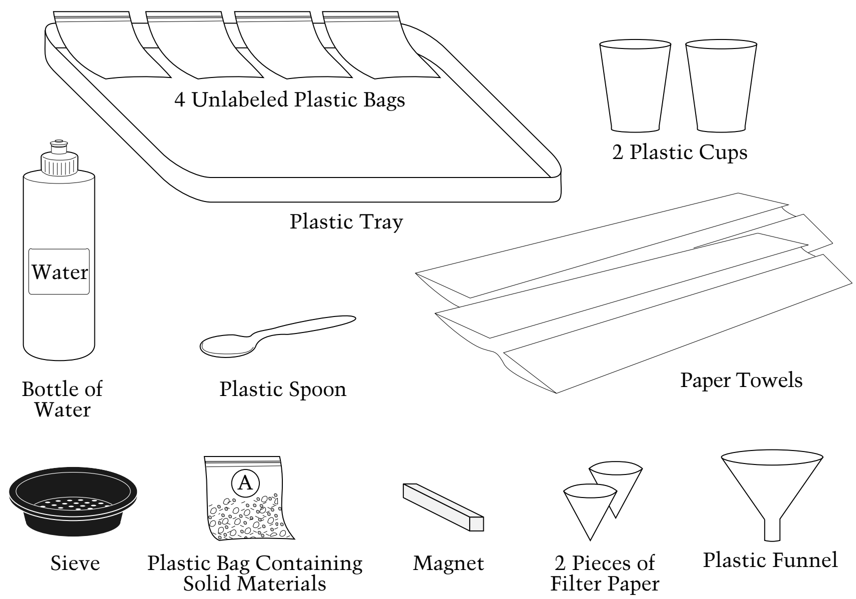
Percentages of Eighth Graders Receiving the Following Scores		
Complete	Essential	Partial
20%	8%	16%

Grade 12: Task Summaries and Sample Questions

SEPARATION

Separating a Mixture of Solid Materials

For this task, you have been given a kit that contains materials that you will use to perform an investigation during the next 30 minutes. Please open your kit now and use the following diagram to check that all of the materials in the diagram are included in your kit. If any materials are missing, raise your hand and the administrator will provide you with the materials that you need.



In this task, students were asked to apply their understanding of basic physical principles and the use of simple laboratory equipment to separate a mixture of five solid materials that have different properties (steel pellets, copper pellets, iron filings, sand, and salt). Students designed the procedures and used them to accomplish the task. This task assessed students' ability to apply their conceptual knowledge of physical principles, to draw inferences from investigative results, and to evaluate and effectively communicate their investigative procedures. It also assessed students' understanding of one aspect of the nature of technology by asking students to apply their knowledge to the design of a practical separation procedure.

Figure 3.4 presents an extended constructed-response question that asks students to state what properties the substances in the mixture have that would allow a magnet, filter paper, and a sieve to be used to separate them. This question was posed at the beginning of the task before students were asked to separate the materials in the mixture using equipment in the kit. Responses were scored according to a four-level guide: *Complete, Essential, Partial,* or *Incorrect.* Also included in Figure 3.4 are sample responses from two students. The first student response correctly identified three properties; however, the second response failed to mention that the magnet would only attract metals containing iron or metals that are magnetic.

Figure 3.4 Sample Question One from the Grade 12 Hands-On Task: Separation

THE NATION'S REPORT CARD naep

Look at the contents of plastic bag (A) without opening it. What properties do the substances in the mixture have that would allow the following equipment to be used to separate the mixture?

Magnet:

Some of the material is magnetic, or iron.

Filter paper:

Some of the mixture looks as though it doesn't dissolve in water.

Sieve:

Some of the mixture is much larger than other parts.

Students' responses were scored using a four-level scoring guide. The first sample response received the highest score, **Complete** because it demonstrated understanding of three properties of the substances that would allow for their separation by a magnet, filter paper, and sieve.

Look at the contents of plastic bag (A) without opening it. What properties do the substances in the mixture have that would allow the following equipment to be used to separate the mixture?

Magnet:

would attract the metal and leave only salt and sand

Filter paper:

would allow only the salt dissolved in the water to filter through.

Sieve:

catch all of the big particles letting the smaller ones slide through.

The second sample response received a score of **Essential** because it demonstrated understanding of two properties of the substances that would allow for their separation by two of the three pieces of equipment.

Percentages of Twelfth Graders Receiving Complete, Essential, and Partial Scores

Complete	2%
Essential	9%
Partial	29%

Figure 3.5 presents an extended constructed-response question that asks students to explain to someone else how to separate the materials in the mixture. Prior to this question students were asked to devise and carry out a methodology to separate the materials. Responses were scored according to a five-level guide: *Complete*, *Essential*, *Adequate*, *Partial*, or *Incorrect*. Also included in Figure 3.5 are sample responses from two students. The first student response received a score of *Complete* because the methodology needed to separate the 5 components of the mixture was clearly explained. The second response described the methodology somewhat clearly, but failed to state after step 1 that the metal balls then had to be separated by the magnet into two components, magnetic and non-magnetic. Thus it received a score of *Essential*.

Figure 3.5 Sample Question Two from the Grade 12 Hands-On Task: Separation

Students' responses were scored using a five-level scoring guide. The first sample response received a score of **Complete** because it described steps that led to the separation of the five components in the mixture.

Based on what you discovered as you worked to separate the materials in the mixture, write in the space below step-by-step instructions that would allow someone else to separate all five solids using the same set of equipment.

Start by sifting the sand, salt, and metal shavings from the metal balls. Separate the balls using the magnet. Then use the magnet to separate the metal shavings from the sand and salt. Then take the sand and salt and place it in a cup along with water and stir until the salt dissolves. Then filter the salt water out of the sand by using the filter paper.

The second sample response received a score of **Essential** because it described how to separate out three of the five components but failed to describe the separation of the steel pellets from the copper pellets.

Based on what you discovered as you worked to separate the materials in the mixture, write in the space below step-by-step instructions that would allow someone else to separate all five solids using the same set of equipment.

① using the sieve, sift out the metal balls, ② pour the contents into the tray and pull the magnet through the contents to attract and separate another metal ③ pour the contents into the funnel with a filter in it, ④ pour water over the material to dissolve the salt ⑤ the sand is left

Percentages of Twelfth Graders Receiving Complete, Essential, Adequate, and Partial Scores

Complete	26%
Essential	32%
Adequate	4%
Partial	21%

Summary

The inclusion of hands-on tasks in the NAEP 1996 science assessment presented a series of operational and analytic challenges as well as opportunities. One of the challenges was to make certain that each kit was produced according to standard specifications. This was possible with most of the equipment. One notable exception was the pencil used in the grade 4 and grade 8 tasks described in this chapter. While the pencils looked the same, their densities varied, thus the measurements taken were somewhat different for each student. Student responses, therefore, had to be assessed on the relative heights of the pencil that floated above the surface of each solution. Another challenge was to make certain that the instructions were clear enough to enable students to work through each hands-on task. Diagrams were incorporated to aid in clarification. For example, in the eighth-grade task, a diagram of the pencil floating in a solution in a graduated cylinder was shown. The diagram also indicated which measurement should be taken.

Scoring of student responses also presented their own unique challenges. Some of the data for the grade 12 task, for example, had to be collected in the field. Students were asked to separate out five components of a mixture and place them in plastic bags. Since water was used in the separation process, certain components were wet when placed in the bags. If the bags were kept for examination at a later date, it would be difficult to ascertain the purity of some of the components because of rusting. Thus the assessment administrators were trained to judge the purity of each component and record this purity on a grid that was printed in each student booklet. These grids were then scored at the same time as the rest of the responses.

Information from the field indicated that students enjoyed having the opportunity to conduct a hands-on task. Fourth-graders, especially, wanted to "do another task." Students and their teachers were also given the opportunity to conduct further activities. Used materials were left in the schools together with a booklet containing activities specifically linked to the materials. Each activity listed a learning goal, the materials required, methodology, and a number of questions devised to act as a platform for discussion. Together, these proved to be an important resource, especially in schools that had little access to equipment.

The responses elicited by the hands-on tasks proved to be a valuable source of information regarding the skills deemed most important in scientific investigation. The results are mixed. In the grade 4 floating pencil task, just over 25 percent of students were able to identify the mystery water and give a correct explanation; however, only 14 percent were able to apply their findings in the hands-on task to a practical situation. In the grade 8 salt solution task, about 40 percent of students measured successfully and close to 60 percent were able to calculate averages; however, it should be remembered that the methodology for calculating an average was given. Graphing and interpolating proved to be more difficult, with 28 percent and 20 percent of students completing each skill successfully. The task at grade 12 was not as structured as those at grades 4 and 8. Students were asked to separate components in a mixture and explain how they did it. Approximately one-quarter of the students wrote instructions clear enough to enable someone else to successfully separate the components.

Three of the ten hands-on tasks administered in the 1996 NAEP science assessment have been described in this chapter. Selected questions from each of these tasks are also included, together with student responses. While these questions give some indication of the skills assessed by each task they by no means show all of them. However, those that are included do give an indication of how students fared on questions relating to skills such as measuring, data collection, and graphing. Further analyses of data stemming from all the hands-on tasks administered in the NAEP 1996 science assessment are planned. These analyses will be discussed in an upcoming report that will also include further discussion of the three hands-on tasks presented in this chapter.

Chapter 4

Exploring a More Inclusive NAEP

The 1996 national and state NAEP science assessments are the first in what is planned to be a series of NAEP science assessments based on the new framework adopted by the National Assessment Governing Board. Because they are based on a new and different framework, results from the NAEP 1996 science assessments are not directly comparable to results from previous NAEP science assessments. The introduction of a new framework, and the concomitant beginning of a new trend line for NAEP science results, affords the program the opportunity to update assessment procedures and instruments, bringing them more in line with current best practice.

One of the areas in which the NAEP program continues to seek improvements is in the inclusion and appropriate assessment of two specific populations: students with disabilities and limited English proficient students. The 1996 national and state NAEP science assessments were conducted using revised criteria for the inclusion of such students, and the results reported in Chapters 1 through 3 of this report are based on samples of students that were administered the assessment using these new criteria for inclusion. The 1996 science assessment also included supplemental samples of schools and students. The supplemental samples were designed to allow the study of the effects of the revised inclusion procedures on assessment results and to investigate the feasibility and impact of further increasing the participation of limited English proficient (LEP) students and students with disabilities by offering assessment accommodations. This chapter describes these additional samples and procedural revisions and presents some initial results on research issues pertinent to the development of a more inclusive NAEP.

Because it serves as the *Nation's Report Card*, the intent of NAEP has been to report results that reflect the achievement of *all* students at a given grade or age. However, practical realities and fiscal constraints have always excluded at least some small percentage of students from the determination of NAEP results. For example, in the most recent NAEP assessments, the small percentage of students who receive home schooling, who attend ungraded schools, who attend special schools for the deaf and blind, or who are incarcerated were not included in the samples because of the logistical challenges and costs associated with identifying and assessing such students.

When reporting on the educational achievement of students in a particular grade, NAEP attempts to include *all* students who are enrolled in that grade at the sampled schools. NAEP samples include students with disabilities (including students who have Individualized

Education Programs (IEPs) or who are receiving special services as a result of section 504 of the Rehabilitation Act) and limited English proficient (LEP) students in approximately the same percentages in which they are found in the general school population. Although NAEP has traditionally included a substantial percentage of these students in its assessment results, the program has always recognized that a subset of a given school's students may not be able to participate in the assessment.

In the past, schools were allowed to exclude students from NAEP for a number of reasons. Some students, such as those with significant cognitive disabilities, did not participate in any large-scale standardized assessments if their teachers judged them to be incapable of such participation. Other students were considered incapable of taking assessments such as NAEP in English. And some students did not participate because NAEP was unable to provide the accommodations that would have made their inclusion possible.

To facilitate the consistent implementation of the program's policies, NAEP has provided specific criteria that staff from the sampled schools (typically the team responsible for the student's IEP or the school staff person most knowledgeable about each student) can use to determine which students should be included in the assessment. By using these criteria, considerable numbers of students with disabilities or LEP students have been assessed. For example, NAEP 1994 results indicate that nearly 13 percent of the nation's fourth graders, 10 percent of the eighth graders, and 8 percent of twelfth graders are classified as students with disabilities or LEP students. More than half of the students with disabilities and LEP students sampled for NAEP (59 percent at fourth grade, 56 percent at eighth grade, and 58 percent at twelfth grade) were assessed as part of the NAEP 1994 assessment. However, the remaining 41 to 44 percent were not assessed.[1]

In recent years, a number of policy, legislative, civil rights, and technical considerations have caused NAEP to look more closely at its administration and assessment procedures and to consider changes that can increase participation among students with disabilities or LEP students.[2] Based on previous studies, [3] as well as recommendations from various offices in the U.S. Department of Education, program procedures have been modified to increase participation among students with disabilities and LEP students. Modifications were made in two areas.[4] First, inclusion criteria for the NAEP 1996 assessment were revised with the intention of making them clearer, more inclusive, and more likely to be applied consistently across jurisdictions participating in the state assessment program. Second, a variety of

---

1 Kaplan, B.A. & Leung, P.T., "Statistical Summary of the 1994 NAEP Samples" in N. Allen, D. Kline, & C. Zelenak, *The NAEP 1994 Technical Report*. (Washington, DC: National Center for Education Statistics, 1996).

2 Olson, J.F. & Goldstein, A.A., "Increasing the Inclusion of Students with Disabilities and Limited English Proficient Students in NAEP." *Focus on NAEP*, 2(1). (Washington, DC: National Center for Education Statistics, 1996).

3 National Academy of Education, *The Trial State Assessment: Prospects and Realities. The Third Report of the National Academy of Education Panel on the Evaluation of the NAEP 1992 Trial State Assessment,* (Stanford, CA: National Academy of Education, 1993).

Ysseldyke, J.E., Thurlow, M.L., McGrew, K.S. & Vanderwood, M., *Making Decisions about the Inclusion of Students with Disabilities in Statewide Assessments* (*Synthesis Report 13*). (Minneapolis, MN: University of Minnesota, National Center on Education Outcomes, 1994).

4 Olson, J.F. & Goldstein, A.A., "Increasing the Inclusion of Students with Disabilities and Limited English Proficient Students in NAEP." *Focus on NAEP*, 2(1). (Washington, DC: National Center for Education Statistics, 1996).

assessment accommodations were offered to students with disabilities whose IEPs specified such accommodations for testing and to LEP students who, in the opinion of their instructors, required an accommodation in order to take the assessment in English.

Several important technical issues need to be solved before the procedural modifications can be implemented as official NAEP policy. One issue is the effect of these modifications on NAEP's capacity to provide accurate comparisons over time. One of NAEP's goals is to report on trends in academic achievement. Accurately reporting changes requires keeping assessment procedures and instrumentation comparable during the period over which measurement is sought. Modifying inclusion criteria and providing accommodations can significantly expand the number of students with disabilities and LEP students included in NAEP assessments. Although this expansion is desirable, it can cloud the interpretation of changes in achievement over time, since assessments conducted using revised procedures might include results for students who would not have been included in previous assessments.

Another issue is the validity of results from nonstandard administrations (i.e., administrations in which accommodations were allowed) and the comparability of these results to results obtained under standard conditions. Specifically, it may not be possible to summarize and report data obtained under nonstandard conditions in terms of the same NAEP scale used for results obtained under standard conditions. The question is, do scale score results obtained under nonstandard conditions convey the same information about educational achievement as corresponding results obtained under standard conditions?

The 1996 national and state assessments in mathematics and science included supplemental samples of schools and students to allow research into inclusion, accommodation, and score validity issues, and to provide a bridge to future mathematics and science assessments in which revised inclusion criteria and the provision of accommodations will be standard program practice.

Based on results from the 1996 national and state NAEP mathematics assessments, initial answers to a number of important research questions were obtained. As discussed in the recently released *NAEP 1996 Mathematics Report Card for the Nation and the States*,[5] the introduction of the revised inclusion criteria, without the provision of accommodations, had little effect on the percentage of the *total* population that was assessed in the NAEP mathematics assessment at either the national or state level and had, at most, a limited effect on the percentage of students with disabilities or LEP students who were assessed. In contrast, the provision of accommodations and adaptations clearly increased participation rates in the mathematics assessment for students with disabilities and LEP students at grades 4 and 8, though not at grade 12. There is evidence at all three grades, however, that some students with disabilities will be assessed with accommodations or adaptations when these are available but will be assessed under standard conditions when special administration procedures are not available. The potential impact of this "switching" phenomenon on trend measurement is a topic for expanded analysis of the mathematics results and discussion in future NAEP reports.

[5] Reese, C.M., Miller, K.E., Mazzeo, J. & Dossey, J.A., *NAEP 1996 Mathematics Report Card for the Nation and the States* (Washington, DC: National Center for Education Statistics, 1997).

The 1996 national and state science results provide further information about NAEP's inclusion-related research questions.

- Results from the grade 8 state NAEP science assessment are consistent with the mathematics results in indicating that the use of the revised inclusion criteria, without the provision of accommodations, had little effect on the overall percentage of the total population assessed, or on the percentages of students with disabilities or LEP students assessed.
- There was some evidence from the national NAEP science assessment that the provision of accommodations resulted in higher rates of participation for students with disabilities and LEP students. However, the effects of providing accommodations were more limited in scope than was observed in the mathematics assessment. The only significant difference in overall population exclusion rates for the science assessment was found at grade 12, and the only significant difference in inclusion rates was observed at grade 4 among students with disabilities. Differences in results between the mathematics and science assessments may be due, at least in part, to differences in the range of adaptations that were offered.
- The national science assessment results at grades 4 and 8 indicate that a portion of the population of students with disabilities were assessed with accommodations when these were available but were assessed under standard conditions when special administration procedures were not available. These results mirror the findings of the 1996 mathematics assessment at these grades and again raise concerns about the impact of this "switching" phenomenon on trend measurement.

The analyses discussed in this chapter, and those reported in the *NAEP 1996 Mathematics Report Card*, are only the first steps in what is an ongoing research and development effort. Whether changes in inclusion and administration procedures affected overall scale score results is a topic for expanded analysis and discussion in future NAEP reports. A comprehensive research report on this and other inclusion issues will be published later in 1997.

The NAEP 1996 National and State Science Samples

The design of the NAEP 1996 science assessment required national samples of schools as well as distinct samples of schools within each jurisdiction that participated in the state assessment program. Three types of samples were drawn.

- In the first type of school sample (denoted S1), the assessment was conducted using the same inclusion criteria used during the 1990 and 1992 NAEP assessments in mathematics. No assessment accommodations were offered to students in S1 schools. As the 1996 NAEP science assessment was based on a new framework, the national NAEP science assessment did not require or include an S1 sample of schools. However, in order to control costs and to minimize burden on jurisdictions participating in state NAEP, state-level samples of science sessions conducted under S1 conditions were obtained. In order to report state-level trend results in mathematics, an S1 sample of schools was required in each jurisdiction. For reasons of efficiency, these same schools also conducted the science assessment.

- In the second type of school sample (denoted S2), revised inclusion criteria were used. S2 samples were drawn for all three grades in the national assessment and for each of the jurisdictions participating in the grade 8 state assessment. No assessment accommodations were offered to students in S2 schools.
- In the third sample (denoted S3), the assessment was conducted using inclusion criteria that were effectively identical to those used in S2 schools. The S3 sample was distinguished, however, by the availability of a variety of assessment accommodations. Because of concerns about feasibility and an interest in managing the burden on participating jurisdictions, separate S3 samples were not obtained for the state assessment.

Thus, the national assessment in science was based on two samples (an S2 and an S3 sample), and the state assessment was based on two distinct samples in each jurisdiction (an S1 and an S2 sample). To ensure sufficient amounts of data for planned analyses, students with disabilities and LEP students were *oversampled* in national S2 and S3 schools and all students in S3 that received an accommodation at a given grade were administered the same NAEP assessment booklet.

Data from the full national S2 sample were analyzed as the official reporting sample for the 1996 national NAEP science assessment. National results reported in Chapters 1 through 3 of this report are based on this sample. For the state assessment, data from the full S2 sample and a portion of the S1 sample (excluding students with IEPs or equivalent plans and students classified as LEP) were combined and analyzed as the reporting samples for jurisdictions participating in state NAEP. State results reported in Chapters 1 through 3 are based on these combined samples. By comparing grade 8 state NAEP science results obtained from the full S1 and full S2 samples, the NAEP program can supplement the information provided by the 1996 mathematics assessment on the effects of changing inclusion criteria on inclusion rates and assessment results. By comparing national results obtained from the full S2 sample to those from the full S3 sample, the program will be able to add to what it has learned from the 1996 mathematics assessment about the effects of providing accommodations.

National and State Percentages of Students with Disabilities and LEP Students

Prior to a NAEP assessment in a school, a staff member designated by the school as the NAEP liaison is presented with a list of sampled students and, in consultation with appropriate school staff, is asked to identify students with disabilities or students classified by the school as LEP. NAEP records this information as part of its standard data collection procedures. Table 4.1 presents the percentages of the national NAEP population at each grade identified as students with disabilities (SD), LEP students, or both.

Table 4.1 — **Percentage of National Population Identified as SD*, LEP**, or Both: Public and Nonpublic Schools Combined**

THE NATION'S REPORT CARD naep

	Total	SD* Only	LEP** Only	Both SD* and LEP**
Grade 4	16	11	4	1
Grade 8	11	9	2	0
Grade 12	8	6	2	0

* Students with Disabilities.
**Limited English Proficient Students.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Twelve percent of the nation's fourth-grade students, 9 percent of the nation's eighth-grade students, and 6 percent of the nation's twelfth-grade students are identified as students with disabilities (i.e., combining "SD Only" and "Both SD and LEP"). Five percent of the nation's fourth graders and 2 percent of the nation's eighth and twelfth graders are identified as LEP students (i.e., combining "LEP Only" and "Both SD and LEP"). These estimates, which are based on the samples of schools and students involved with the NAEP science assessment, closely mirror the corresponding estimates obtained from the NAEP 1996 mathematics assessment.

Analogous results for grade 8 public schools are provided in Appendix D (Table D.1) for the nation and for each of the jurisdictions participating in the state assessment.[7] The results are consistent with those obtained from the mathematics assessment in indicating substantial variation across states and jurisdictions in the percentages of students with disabilities and LEP students. (See Appendix D for further discussion.)

[7] Throughout this chapter, results from the state assessment are limited to public school students. State-level samples of nonpublic school students were relatively modest in size and, for a substantial number of jurisdictions, did not meet minimum NCES participation rate guidelines established for the reporting of results (see Appendix A). Hence, they were excluded from this chapter and from Appendix D in the interests of clarity and brevity.

Revisions to the Inclusion Criteria

Revised inclusion criteria for NAEP were implemented in the S2 and S3 samples for the 1996 assessment. The revision had four goals:

1. To increase inclusion rates for students with disabilities
2. To bring NAEP inclusion rules for LEP students more in line with those used in state testing programs
3. To allow for more consistent inclusion decisions across states and jurisdictions
4. To ensure that inclusion decisions were related to the subject-matter instruction given to the student rather than less relevant considerations

The original inclusion criteria (used in S1 state-level samples) provided a basis for determining whether students could be *excluded* from the assessment. Based on the S1 criteria (i.e., the criteria used in NAEP's mathematics assessments in 1990 and 1992), students with disabilities *could be excluded* only if they were mainstreamed in academic subjects less than 50 percent of the time AND/OR judged to be incapable of participating meaningfully in the assessment. LEP students *could be excluded* if they were native speakers of a language other than English AND enrolled in an English-speaking school for less than two years AND judged to be incapable of taking part in the assessment.

The guidelines used in the S2 samples were revised to emphasize criteria for the *inclusion* rather than exclusion of students with disabilities and LEP students. Although the original criteria did instruct school staff, when in doubt, to include students, the revised criteria were designed to communicate more clearly a presumption of inclusion except under special circumstances. Students with IEPs were to be *included* in the NAEP assessment except in the following cases:

1. The school's IEP team determined that the student could not participate; OR
2. The student's cognitive functioning was so severely impaired that she or he could not participate; OR
3. The student's IEP required that the student be tested with an accommodation or adaptation and that the student could not demonstrate his or her knowledge without that accommodation.

Under the revised criteria, all LEP students receiving academic instruction in English for three years or more were to be included in the assessment. Those LEP students receiving instruction in English for less than three years were to be *included* unless school staff judged them as being incapable of participating in the assessment in English.

In the S3 sample, the revised criteria were used and various accommodations were made available. NAEP attempted to assess students with disabilities under conditions identical to those under which they normally participate in large-scale assessments. To the extent possible, NAEP offered S3 students the assessment accommodations that were specified in their IEP or equivalent document. For example, if a student's IEP specified that he or she could only be assessed with extended assessment time, NAEP provided this accommodation. Thus, students whose IEPs required accommodations were included in NAEP if the program was able to offer their accommodation.

Accommodations Provided

An array of assessment accommodations were permitted. In general, most accommodations that schools routinely provided for their own testing were allowed in S3. These permitted accommodations included:

- One-on-one testing
- Small group testing
- Extended time
- Oral reading of directions
- Signing of directions
- Use of magnifying equipment
- Use of an individual to record answers

NAEP also allowed the school to use photocopy equipment to produce enlarged versions of test booklets for students with visual disabilities. Enlarged booklets were made available to students who normally would have been assessed using large-print materials.

It should be noted that students assessed under one of the special conditions typically received some combination of accommodations. For example, students assessed in small groups (as opposed to standard NAEP sessions of roughly 30 students) usually received extended time and had directions and/or assessment questions read aloud as needed. In one-on-one administrations, students often received assistance in recording answers, had directions and questions read aloud, and were afforded extra time.

NAEP goals and plans regarding LEP students were somewhat different. As with students with disabilities, the new inclusion criteria emphasized inclusion rather than exclusion, and LEP students were eligible for any of the accommodations previously listed. However, field test experience had suggested that many LEP students do not have IEPs that specify assessment accommodations. Because the majority of these students are native Spanish speakers, a Spanish/English glossary of scientific terms used in the assessment was produced. This glossary was made available to students at all three grades who, when tested, normally make use of such a document or who typically receive related accommodations (such as a bilingual dictionary). Use of the glossary was permitted in standard NAEP testing sessions, as well as in small group and one-on-one testing situations. Students using the glossary were typically given extra time.

State NAEP Science Results on the Effects of Revised Inclusion Criteria

State-by-state percentages for the S1 and S2 samples for grade 8 public school students are presented in Appendix D (Table D.2). Results from the state NAEP science assessment were consistent with those from the state and national NAEP mathematics assessments in indicating that the revised inclusion criteria, without the provision of accommodations, had little effect on the percentage of the population assessed in NAEP. A significant difference in exclusion percentages between the S1 and S2 samples was found in only one of 43 jurisdictions. This same jurisdiction (Delaware) exhibited a significant difference in grade 8 exclusion rates for the mathematics assessment. Averaged across jurisdictions, there was virtually no difference between the S1 and S2 samples in the percentage of the grade 8 population excluded from NAEP.

At the national level, and in most of the jurisdictions that participated in the state assessment, students with disabilities and LEP students constituted a relatively modest percentage of the total school population. Because the effects of the inclusion criteria were confined to these groups, examining exclusion rates within the total population may not provide a sufficiently sensitive measure of their effects. Examining inclusion rates among students with disabilities and LEP students provides a more in-depth analysis and affords a potentially different perspective on the procedural changes.

Appendix D (Tables D.3 and D.4) presents the percentages of students with disabilities and LEP students who were assessed in each of the jurisdictions that participated in the grade 8 state NAEP science assessment. State results again show no consistent pattern of increased inclusion rates among students with disabilities or LEP students. A statistically significant difference in inclusion rates for students with disabilities was found only in Delaware. Only four jurisdictions had sufficiently large samples of LEP students to support comparisons of S1 and S2 inclusion rates. Although observed inclusion rates for all four jurisdictions were higher in the S2 samples than in the S1 samples, none of these differences were statistically significant.

National NAEP Science Results on the Effects of Providing Accommodations

As noted earlier, comparisons of national science results from the S3 sample with those obtained from the S2 sample help to assess the effects of providing accommodations. Results from the mathematics assessment indicated that the provision of accommodations and adaptations had only modest effects on exclusion rates, with a significant decrease obtained at only one of the three grades. The science results shown in Table 4.2 are similar to the mathematics results in some ways but different in others. Although observed national exclusion rates for the science assessment were lower in the S3 sample than in the S2 sample for grades 4 and 12, the difference was significant only at grade 12. The grade 12 difference, though small, corresponds to a 36 percent reduction in the percentage of the population that is excluded. The presence of a significant effect at grade 12 in the science results is somewhat different than the mathematics results where the only significant reduction occurred at grade 4, not grade 12.

Table 4.2 Percentage of National Population Excluded From the Assessment: Public and Nonpublic Schools Combined

THE NATION'S REPORT CARD naep

	S2: Using Revised Inclusion Criteria	S3: Using Revised Criteria and Providing Accommodations
Grade 4	8	6
Grade 8	4	4
Grade 12	4	3 †

† Indicates a significant difference between S2 and S3 results.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

In order to more directly examine the effects of providing accommodations on the populations of students to which they were targeted, Table 4.3 presents national percentages of students with disabilities and LEP students assessed under standard conditions, the percentages assessed with the provision of accommodations, and the total percentages of these students who were assessed. Results from the NAEP mathematics assessment clearly indicated an increase in inclusion percentages for students with disabilities and LEP students at grades 4 and 8. The NAEP science results, though consistent in some respects, provide less convincing evidence of such an increase. At all three grades, the observed percentages of students with disabilities and LEP students were higher in the S3 samples than in the S2 samples. However, with one exception, the observed differences in inclusion rates were not statistically significant.

Table 4.3

Percentage of Students with Disabilities and Limited English Proficient Students in the National Population Included in the Assessment: Public and Nonpublic Schools Combined

	SD*		LEP**	
	S2: Using Revised Inclusion Criteria	S3: Using Revised Criteria and Providing Accommodations	S2: Using Revised Inclusion Criteria	S3: Using Revised Criteria and Providing Accommodations
Grade 4				
Assessed Under Standard Conditions	48	35†	43	42
Assessed With Accommodations		34		11
Total Assessed	48	70†	43	53
Grade 8				
Assessed Under Standard Conditions	61	40†	62	50
Assessed With Accommodations		27		14
Total Assessed	61	67	62	64
Grade 12				
Assessed Under Standard Conditions	48	38	66	80
Assessed With Accommodations		18		3
Total Assessed	48	56	66	83

* Students with Disabilities.
** Limited English Proficient Students.
† Indicates a significant difference between S2 and S3 results.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Among students with disabilities, the grade 4 difference between S3 and S2 inclusion rates is statistically significant. Seventy percent of grade 4 students with disabilities were assessed in the S3 sample compared to 48 percent in the S2 sample, a 22 percent difference. Observed differences at the other two grades were 7 and 8 percent, respectively. Neither of these differences was statistically significant. Among LEP students, observed differences in inclusion rates ranged from 17 percent (at grade 12) to 2 percent (at grade 8). However, even the largest of these differences was not statistically significant.

It should be noted here that the science findings on inclusion for LEP students run somewhat counter to those reported in the *1996 Mathematics Report Card*. In the NAEP mathematics assessment, the provision of accommodations and adaptations resulted in significantly greater numbers of LEP students being included in the assessment at grades 4 and 8. No significant differences were obtained in the science assessment. The absence of a significant result in science may be due, at least in part, to differences in the array of accommodations and adaptations offered. The 1996 mathematics assessment included a Spanish-bilingual version of a NAEP assessment booklet at grades 4 and 8. Use of this booklet was the most frequently employed accommodation/adaptation among LEP students at these grades. Due to resource constraints, and to lingering questions about the comparability of results obtained with standard and translated versions of NAEP instruments, bilingual versions of the NAEP science assessment were not developed or offered. Instead, students were offered a Spanish/English glossary of scientific terms used in the assessment. Very few of the LEP students in the science assessment made use of this accommodation. Thus, the absence of a bilingual booklet and the lack of use of the Spanish/English glossary may partly account for the lack of a statistically significant increase in LEP inclusion rates. However, it is also worth noting that observed inclusion rates for LEP students were higher in the S3 sample than in S2 sample for all three grades. The absence of statistically significant effects may also partly be the result of the relatively small numbers of such students encountered in NAEP samples.

As discussed above, expanded inclusion for students with disabilities and LEP students is desirable, but it presents challenges for the measurement of trends. Changes in overall rates of exclusion present one such challenge. The overall exclusion rate data presented in Table 4.2 suggest that such changes are small and perhaps can be ignored when measuring trends. This issue will be analyzed and discussed in greater detail in forthcoming NAEP reports. Additional challenges to trend measurement are associated with the availability of accommodations. In any population of students with disabilities or LEP students, some students may be capable of taking the assessment under standard conditions, but they may do somewhat better or be more comfortable with an accommodation. Results obtained with accommodations may be more valid, particularly from the perspective of the individual student. However, assessing such students without the benefit of accommodations in one assessment and providing such accommodations in a later assessment can complicate the interpretation of trend results.

NAEP science results in Table 4.3 suggest that some students with disabilities will be assessed with accommodations when they are available but will be assessed under standard conditions when special administration procedures are not available. This finding is consistent with results from the mathematics assessment. At all three grades in the 1996 science assessment, the observed percentage of students with disabilities who were assessed without accommodations was lower in the S3 sample than in the S2 sample. For example, the percentages were 13 percent lower in S3 than in S2 at grade 4, and 21 percent lower at grade 8. Both the grade 4 and grade 8 differences are statistically significant. At grade 12, the observed S3 inclusion percentage is 10 points lower than that obtained in the S2 sample, but this difference is not statistically significant.

This phenomenon was not evident among LEP students. There is no consistent pattern of results indicating that fewer LEP students were assessed under standard conditions than when accommodations were present. At grade 4 there was almost no difference between the S2 and S3 samples in the observed percentages of LEP students assessed under standard conditions. At grade 8, the observed percentage was 12 percent lower in the S3 sample than in the S2 sample while at grade 12, the percentage was 14 percent higher in the S3 sample than in the S2 sample. At all grades, these observed differences were not statistically significant.

The potential effect on trend measurement of this "switching" phenomenon in the students with disabilities population, and its absence in the LEP student population, are additional topics for expanded analysis and discussion in future NAEP reports.

Concluding Comments

Increasing the numbers of students with disabilities and LEP students who meaningfully participate in the NAEP assessment remains an important program goal. To the extent possible, NAEP results should represent the performance of *all* students. Greater inclusiveness in a nationally visible program like NAEP emphasizes that *all* students, including those with special needs, are entitled to a quality education and that we, as a nation, care about the educational achievement of all our students.

The NAEP program benefits from greater inclusiveness in other ways, as well. Other things being equal, greater inclusiveness improves NAEP's validity because achievement comparisons across assessment years, or across jurisdictions participating in the state assessment, can be made with greater confidence. However, increasing the participation of students with disabilities and LEP students can ideally be accomplished in a manner that does not jeopardize the program's ability to meet another important goal — the measurement of educational progress over time. The results described in this chapter and those reported in Chapter 4 of the *1996 Mathematics Report Card* were made possible by embedding an experimental design within the NAEP 1996 assessment. This experiment allowed the program to study the impacts of proposed procedural changes on important program results, such as inclusion rates and estimates of achievement, and provides a bridge to future assessments in which the proposed procedural changes will be standard NAEP policy.

Although they provide useful information, the analyses discussed in this chapter are only the first steps in an ongoing research and development effort. Additional questions remain about the validity of results when accommodations are used and about the comparability of these results to results obtained under standard conditions. The impact of providing accommodations on NAEP estimates of scale score distributions, for the total population and for some of NAEP's traditional reporting subgroups (e.g., Black and Hispanic students), also requires further study. In-depth analyses of the data gathered with NAEP's SD/LEP Questionnaires can provide more detailed information about the nature and extent of student disabilities, the exposure of these students to appropriate grade-level curriculum, the assessment practices that schools use with these students, and the nature of the students excluded from NAEP assessments. Analyses pertinent to these and other research issues will be included in future NAEP reports.

Appendix A

Overview of Procedures Used for the NAEP 1996 Science Assessment

Introduction

Conducting a large-scale assessment such as the National Assessment of Educational Progress (NAEP) entails the successful coordination of numerous projects, committees, procedures, and tasks. This appendix provides an overview of the NAEP 1996 science assessment's primary components: framework, instrument development, administration, scoring, and analysis. A more extensive review of the procedures and methods used in the science assessment will be included in two subsequent technical reports: *NAEP 1996 Technical Report* and *Technical Report of the NAEP 1996 State Assessment Program in Science.*

The NAEP 1996 Science Assessment

The science framework for the 1996 National Assessment of Educational Progress was produced under the auspices of the National Assessment Governing Board through a consensus process managed by the Council of Chief State School Officers, who worked with the National Center for Improving Science Education and the American Institutes for Research. The framework was developed over a ten-month period between October 1990 and August 1991. The following factors guided the process for developing consensus on the science framework:[1]

- The active participation of individuals such as curriculum specialists, science teachers, science supervisors, state assessment developers, administrators, individuals from business and industry, government officials, and parents;
- The representation of what is considered essential learning in science, and the recommendation of innovative assessment techniques to probe the critical abilities and content areas; and
- The recognition of the lack of agreement on a common scope of instruction and sequence, components of scientific literacy, important outcomes of learning, and the nature of overarching themes in science.

[1] *Science Framework for the 1996 National Assessment of Educational Progress* (Washington, DC: National Assessment Governing Board, 1994).

While maintaining some conceptual continuity with the NAEP 1990 science assessment, the 1996 framework takes into account the current reforms in science education, as well as documents such as the science framework used for the 1991 International Assessment of Educational Progress. In addition, the Framework Steering Committee recommended that a variety of strategies be used for assessing students' performance. These included:

- Performance tasks that allow students to manipulate physical objects and draw scientific understanding from the materials before them,
- Constructed-response questions that provide insights into students' levels of understanding and ability to communicate in the sciences as well as their ability to generate, rather than simply recognize, information related to scientific concepts and their interconnections, and
- Multiple-choice questions that probe students' conceptual understanding and ability to connect ideas in a scientifically sound way.

Percentage of Assessment Time by Major Dimensions of Framework

The framework for the 1996 science assessment is represented as a matrix with two dimensions represented by three fields of science (earth, physical, and life) and three elements of knowing and doing science (conceptual understanding, scientific investigation, and practical reasoning). In addition, there are two overarching domains that describe science, nature of science and themes. Figures A.1a, A.1b, and A.1c describe, respectively, the fields of science, the elements of knowing and doing, and the overarching domains that guided the development of the 1996 science assessment.

Figure A.1a — Descriptions of the Three Fields of Science

THE NATION'S REPORT CARD naep

Earth Science

The earth science content assessed centers on objects and events that are relatively accessible or visible. The concepts and topics covered are solid Earth (lithosphere), water (hydrosphere), air (atmosphere), and the Earth in space. The solid Earth consists of composition; forces that alter its surface; the formation, characteristics and uses of rocks; the changes and uses of soil; natural resources used by humankind; and natural forces within the Earth. Concepts and topics related to water consist of the water cycle; the nature of oceans and their effects on water and climate; and the location of water, its distribution, characteristics, and effect of and influence on human activity. The air is broken down into composition and structure of the atmosphere (including energy transfer); the nature of weather; common weather hazards; and air quality and climate. The Earth in space consists of setting of the Earth in the solar system; the setting and evolution of the solar system in the universe; tools and technology that are used to gather information about space; apparent daily motions of the Sun, the Moon, the planets and the stars; rotation of the Earth about its axis, and the Earth's revolution around the Sun; and tilt of the Earth's axis that produces seasonal variations in the climate.

Physical Science

The physical science component relates to basic knowledge and understanding concerning the structure of the universe as well as the physical principles that operate within it. The major sub-topics probed are matter and its transformations, energy and its transformations, and the motion of things. Matter and its transformations are described by diversity of materials (classification and types and the particulate nature of matter); temperature and states of matter; properties and uses of material (modifying properties, synthesis of materials with new properties); and resource management. Energy and its transformations involve different forms of energy; energy transformations in living systems, natural physical systems, and artificial systems constructed by humans; and energy sources and use, including distribution, energy conversion, and energy costs and depletion. Motion is broken down into an understanding of frames of reference; force and changes in position and motion; action and reaction; vibrations and waves as motion; general wave behavior; electromagnetic radiation; and the interactions of electromagnetic radiation with matter.

Life Science

The fundamental goal of life science is to attempt to understand and explain the nature and function of living things. The major concepts assessed in life science are change and evolution, cells and their functions (not at grade 4), organisms, and ecology. Change and evolution includes diversity of life on Earth; genetic variation within a species; theories of adaptation and natural selection; and changes in diversity over time. Cells and their functions consists of information transfer; energy transfer for the construction of proteins; and communication among cells. Organisms are described by reproduction, growth and development; life cycles; and functions and interactions of systems within organisms. The topic of ecology centers on the interdependence of life—populations, communities, and ecosystems.

SOURCE: *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1995).

Figure A.1b Descriptions of Knowing and Doing Science

THE NATION'S REPORT CARD naep

Conceptual Understanding

Conceptual understanding includes the body of scientific knowledge that students draw upon when conducting a scientific investigation or engaging in practical reasoning. Essential scientific concepts involve a variety of information including facts and events the student learns from science instruction and experiences with the natural environment and scientific concepts, principles, laws, and theories that scientists use to explain and predict observations of the natural world.

Scientific Investigation

Scientific investigation probes students' abilities to use the tools of science, including both cognitive and laboratory tools. Students should be able to acquire new information, plan appropriate investigations, use a variety of scientific tools, and communicate the results of their investigations.

Practical Reasoning

Practical reasoning probes students' ability to use and apply science understanding in new, real-world applications.

SOURCE: *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1995).

Figure A.1c Description of Overarching Domains

THE NATION'S REPORT CARD naep

The Nature of Science

The nature of science incorporates the historical development of science and technology, the habits of mind that characterize these fields, and methods of inquiry and problem-solving. It also encompasses the nature of technology that includes issues of design, application of science to real-world problems, and trade-offs or compromises that need to be made.

Themes

Themes are the "big ideas" of science that transcend the various scientific disciplines and enable students to consider problems with global implications. The NAEP science assessment focuses on three themes: systems, models, and patterns of change.

- Systems are complete, predictable cycles, structures or processes occurring in natural phenomena. Students should understand that a system is an artificial construction created to represent, or explain a natural occurrence. Students should be able to identify and define the system boundaries, identify the components and their interrelationships and note the inputs and outputs to the system.
- Models of objects and events in nature are ways to understand complex or abstract phenomena. As such they have limits and involve simplifying assumptions but also possess generalizability and often predictive power. Students need to be able to distinguish the idealized model from the phenomenon itself and to understand the limitations and simplified assumptions that underlie scientific models.
- Patterns of change involve students' recognition of patterns of similarity and differences, and recognize how these patterns change over time. In addition, students should have a store of common types of patterns and transfer their understanding of a familiar pattern of change to a new and unfamiliar one.

SOURCE: *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1995).

Table A.1 summarizes the distribution of assessment time across the three fields of science — earth, physical, and life. These fields provide the basis for the content area scales. Care was taken to ensure congruence between the percentages used in the assessment (actual) and those indicated in the assessment specifications (recommended).

Table A.1 **Distribution of Assessment Time by Field of Science**

	Earth		Physical		Life	
	Actual	Recommended*	Actual	Recommended*	Actual	Recommended*
Grade 4	33%	33%	34%	33%	33%	33%
Grade 8	30%	30%	30%	30%	40%	40%
Grade 12	33%	33%	33%	33%	34%	33%

* *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1995).

Table A.2 shows the distribution of assessment time across the second dimension, knowing and doing science. This dimension includes conceptual understanding, scientific investigation, and practical reasoning.

Table A.2 **Distribution of Assessment Time by Knowing and Doing Science**

	Conceptual Understanding		Scientific Investigation		Practical Reasoning	
	Actual	Recommended*	Actual	Recommended*	Actual	Recommended*
Grade 4	45%	45%	38%	45%	17%	10%
Grade 8	45%	45%	29%	30%	26%	25%
Grade 12	44%	45%	28%	30%	28%	25%

* *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1995).

A number of questions that assess each of the fields of science and each of the ways of knowing and doing science also probe the nature of science and themes (systems, models, and patterns of change). Tables A.3 and A.4 show the actual and recommended percentages of assessment time for these two overarching domains.

Table A.3 Distribution of Assessment Time for Nature of Science

THE NATION'S REPORT CARD naep

	Actual	Recommended*
Grade 4	19%	≥15%
Grade 8	21%	≥15%
Grade 12	31%	≥15%

* *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1995).

Table A.4 Distribution of Assessment Time for Themes

THE NATION'S REPORT CARD naep

	Actual	Recommended*
Grade 4	53%**	33%
Grade 8	49%	50%
Grade 12	55%	50%

* *Science Framework for the 1996 National Assessment of Educational Progress.* (Washington, DC: National Assessment Governing Board, 1995).

** Several of the hands-on tasks were classified as themes

The Assessment Design

Each student in the assessment received a booklet containing six sections. Three of these sections were blocks[2] of cognitive questions that assessed the knowledge and skills outlined in the framework. The other three sections were sets of background questions. Two of the three cognitive sections contained only paper-and-pencil questions, and the third section consisted of a hands-on task with related paper-and-pencil questions. Students at grades 8 and 12 were allowed 30 minutes to complete each cognitive section, while students at grade 4 were given cognitive blocks that required only 20 minutes to complete.

At each grade level there were 15 different sections or blocks of cognitive questions usually consisting of both multiple-choice and constructed-response questions. Each student's booklet contained three of these blocks of cognitive questions. Short constructed-response questions required a few words or a sentence or two for an answer (e.g., briefly stating how nutrients move from the digestive system to the tissues) while extended constructed-response questions generally required a paragraph or more (e.g., outlining an experiment to test the effect of increasing the amount of available food on the rate of increase of the hydra population). Some extended constructed-response questions also required diagrams, graphs, or calculations. It was expected that students could adequately answer the short constructed-response questions in about two to three minutes and the extended constructed-response questions in about five minutes.

Other features were built into the blocks of questions. Four of the blocks at each grade level were hands-on tasks where students were given a set of equipment and asked to conduct an investigation and answer questions relating to the investigation. Every student conducted a hands-on task that was always presented as the third cognitive section. A second feature was the inclusion of theme blocks at each grade level — one assessing systems, one assessing models, and one assessing patterns of change. For example, students were shown a simplified model of part of the solar system with a brief description, and then asked a number of questions based on this scenario. Theme blocks were placed randomly in the student booklets, but not in every booklet. No student received more than one theme block.

Each booklet in the assessment also included three sets of student background questions. The first, consisting of general background questions, asked students about their race/ethnicity, mother's and father's level of education, reading materials in the home, homework, school attendance, and, at grade 12, academic expectations. The second, consisting of science background questions, asked students questions about their classroom learning activities such as hands-on exercises, courses taken, use of specialized resources such as computers, and views on the utility and value of science. (Students were given five minutes to complete each of these sets of questions, with the exception of the fourth graders, who were given more time because the general background questions were read aloud to them.) The third set contained five questions about students' motivation to do well on the assessment, their perception of the difficulty of the assessment, and their familiarity with the types of cognitive questions asked; this section took three minutes or less to complete.

[2] "Blocks" are collections of questions grouped, in part, according to the amount of time required to answer them.

The data in Table A.5 reflect the number of questions by type and by grade level for the 1996 assessment. The assessment pool contained 443 unique questions — 165 multiple-choice (MC), 219 short constructed-response (SCR), and 59 extended constructed-response (ECR). Some of these questions were used at more than one grade level; thus, the sum of the questions that appear at each grade level is greater than the total number of unique questions.

Table A.5 **Distribution of Questions by Type**

THE NATION'S REPORT CARD naep

	Grade 4			Grade 8			Grade 12		
	MC[1]	SCR[2]	ECR[3]	MC[1]	SCR[2]	ECR[3]	MC[1]	SCR[2]	ECR[3]
Grade 4 Only	42	57	12						
Grades 4 & 8 Overlap	9	16	4	9	16	4			
Grade 8 Only				44	58	13			
Grades 8 & 12 Overlap				21	26	3	21	26	3
Grade 12 Only							49	62	27
TOTAL by Grade	51	73	16	74	100	20	70	88	30

[1] Multiple-choice questions

[2] Short constructed-response questions

[3] Extended constructed-response questions

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Using information gathered from the field test, the booklets were carefully constructed to balance time requirements for the question types in each block. More information on the design of the assessment is presented in the forthcoming *NAEP 1996 Technical Report*.

In addition to the student assessment booklets, three other instruments provided data relating to the assessment: a teacher questionnaire, a school characteristics and policy questionnaire, and a questionnaire designed to gather information about students with disabilities (SD) and limited English proficient (LEP) students.

The teacher questionnaire was administered to the science teachers of the fourth- and eighth-grade students participating in the assessment. The questionnaire consisted of three sections and took approximately 20 minutes to complete. The first section focused on the teacher's general background and experience; the second section focused on the teacher's background related to science; and the third section focused on classroom information about science instruction.

The school characteristics and policy questionnaire was given to the principal or other administrator in each participating school and took about 20 minutes to complete. The questions asked about school policies, programs, facilities, and the demographic composition and background of the students.

The SD/LEP student questionnaire was completed by a school staff member knowledgeable about those students who were selected to participate in the assessment and who were identified as (1) having an Individualized Education Plan (IEP) or equivalent plan (for reasons other than being gifted or talented) or (2) being limited English proficient (LEP). A questionnaire was completed for each SD/LEP student sampled regardless of whether the student participated in the assessment. Each questionnaire took approximately five minutes to complete and asked about the student and the special programs in which he or she participated.

National and State Samples

The national and regional results presented in this report are based on nationally representative probability samples of fourth-, eighth-, and twelfth-grade students. The samples were selected using a complex multistage sampling design that involved sampling students from selected schools within selected geographic areas across the country. The sample design had the following stages:

1. Selection of geographic areas (a county, group of counties, or metropolitan statistical area)
2. Selection of schools (public and nonpublic) within the selected areas
3. Selection of students within the selected schools

Each selected school that participated in the assessment and each student assessed represents a portion of the population of interest. Sampling weights are needed to make valid inferences between the student samples and the respective populations from which they were drawn. In addition, NAEP oversamples nonpublic schools and schools in which more than 15 percent of the student population is non-White. Sampling weights adjust for disproportionate representation due to such oversampling.

Table A.6 provides a summary of the weighted and unweighted student sample sizes for the national NAEP 1996 science assessment. The numbers reported include public and nonpublic school students.

Table A.6 National School and Student Sample Sizes for the NAEP 1996 Science Assessment

THE NATION'S REPORT CARD naep

	Number of Schools	Unweighted Student Sample Size	Weighted Student Sample Size
Grade 4	237	7,305	3,621,677
Grade 8	202	7,774	3,568,034
Grade 12	232	7,537	2,907,065

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

The results of the 1996 state assessment program in science provided in this report are based on state-level samples of eighth-grade students. The samples of both public and nonpublic school eighth-grade students were selected based on a two-stage sample design that entailed selecting schools within participating jurisdictions and selecting students within schools. The first-stage samples of schools were selected with a probability proportional to the eighth-grade enrollment in those schools. Special procedures were used for jurisdictions that had many small schools and for jurisdictions that had a small number of schools. In addition, each jurisdiction was provided with a list of substitute schools. For each sampled school, a substitute school was designated that was matched as closely as possible to the characteristics of the sampled school. States were permitted to replace a sampled school that refused to participate with its designated substitute school.

As with the national samples, the jurisdiction samples were weighted to allow for valid inferences about the populations of interest. Tables A.7a and A.7b contain, for public and nonpublic schools respectively, the unweighted numbers of participating schools and students as well as weighted school and student participation rates. Two weighted school participation rates are provided for each jurisdiction. The first rate is the weighted percentage of schools participating in the assessment before substitution. This rate is based only on the number of schools that were initially selected for the assessment. The numerator of this rate is the sum of the number of students represented by each initially selected school that participated in the assessment. The denominator is the sum of the number of students represented by each of the initially selected schools that had eligible students enrolled. This rate included both participating and nonparticipating schools.

The second school participation rate is the weighted participation rate after substitution. The numerator of this rate is the sum of the number of students represented by each of the participating schools, whether originally selected or substituted. The denominator is the same as that for the weighted participation rate for the initial sample. This statement means that for a given jurisdiction, the weighted participation rate after substitution is at least as great as the weighted participation rate before substitution.

Also presented in Tables A.7a and A.7b are the weighted percentages of students who participated after make-up sessions were completed. This rate reflects the percentage of the eligible student population from participating schools within the jurisdiction, and this percentage represents the students who participated in the assessment in either an initial session or a make-up session. The numerator of this rate is the sum, across all assessed students, of the number of students represented by each selected student who was eligible to participate, including students who did not participate.

Table A.7a

NAEP 1996 School and Student Participation Rates by Jurisdiction: Grade 8, Public Schools

THE NATION'S REPORT CARD naep

	Weighted School Participation Rate		Total Number of Schools Participating	Weighted Student Participation Rate	Total Number of Students Assessed
	Before Substitutes	After Substitutes			
Nation	**80**	**80**	**128**	**93**	**6,376**
Alabama	84	90	96	93	2,186
Alaska ‡	93	93	55	82	1,517
Arizona	87	87	94	90	2,151
Arkansas ‡	70	71	76	92	1,858
California	83	94	101	92	2,292
Colorado	100	100	108	91	2,514
Connecticut	100	100	102	93	2,489
Delaware	100	100	30	89	1,903
District of Columbia	100	100	33	85	1,700
Florida	100	100	105	90	2,353
Georgia	99	99	100	92	2,470
Hawaii	100	100	51	90	2,153
Indiana	87	90	96	92	2,313
Iowa ‡	73	83	91	94	2,172
Kentucky	87	92	100	94	2,459
Louisiana	100	100	111	90	2,615
Maine	91	91	95	92	2,254
Maryland ‡	86	86	89	89	2,092
Massachusetts	92	92	98	91	2,287
Michigan ‡	70	87	92	90	2,186
Minnesota	86	88	95	92	2,383
Mississippi	89	95	103	92	2,469
Missouri	93	96	105	92	2,389
Montana ‡	70	76	79	92	2,029
Nebraska	99	100	120	92	2,724
Nevada ‡	37	38	28	92	964
New Hampshire ‡	66	68	64	90	1,710
New Jersey ‡	63	64	67	93	1,573
New Mexico	100	100	90	90	2,377
New York ‡	70	78	82	90	1,876
North Carolina	100	100	107	91	2,616
North Dakota	80	93	108	94	2,489
Oregon	86	92	100	89	2,275
Rhode Island	90	90	43	89	2,087
South Carolina ‡	86	87	91	90	2,162
Tennessee	92	92	99	91	2,287
Texas	91	96	102	92	2,300
Utah	100	100	94	90	2,715
Vermont ‡	74	75	78	93	1,914
Virginia	100	100	106	90	2,552
Washington	94	95	105	90	2,501
West Virginia	100	100	105	93	2,602
Wisconsin ‡	78	78	90	90	2,148
Wyoming	100	100	67	93	2,619
DDESS	100	100	11	95	602
DoDDS	100	100	58	93	2,223
Guam	100	100	6	90	930

National results are based on the national assessment sample, not on aggregated state assessment program samples.

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public school participation rates (see Appendix A).

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools

DoDDS: Department of Defense Dependents Schools (Overseas)

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table A.7b

NAEP 1996 School and Student Participation Rates by Jurisdiction: Grade 8, Nonpublic Schools

THE NATION'S REPORT CARD naep

	Weighted School Participation Rate		Total Number of Schools Participating	Weighted Student Participation Rate	Total Number of Students Assessed
	Before Substitutes	After Substitutes			
Nation	**77**	**77**	**81**	**97**	**1,398**
Alabama ‡	60	60	10	95	144
Arkansas ‡	74	74	6	99	89
California ‡	80	80	14	96	206
Connecticut ‡	63	65	20	96	263
Delaware ‡	42	44	13	96	313
District of Columbia ‡	52	52	19	95	259
Georgia	88	88	9	96	232
Iowa	94	94	14	96	246
Kentucky ‡	82	82	13	97	260
Louisiana ‡	75	75	21	96	424
Maryland ‡	61	64	19	94	322
Massachusetts ‡	75	77	21	94	335
Michigan ‡	80	87	21	97	332
Minnesota ‡	84	84	19	94	247
Missouri	94	100	24	95	365
Montana	93	97	13	93	154
Nebraska ‡	78	84	20	96	333
Nevada	90	90	8	91	133
New Hampshire ‡	83	83	12	95	179
New Jersey ‡	62	64	20	96	287
New Mexico	95	95	13	95	230
New York ‡	84	87	28	97	514
North Dakota ‡	70	78	10	93	160
Oregon ‡	26	26	4	86	54
Rhode Island ‡	68	68	22	96	340
South Carolina ‡	69	69	8	95	138
Texas ‡	79	79	7	98	130
Utah ‡	64	64	4	93	93
Vermont ‡	72	80	10	91	115
Washington	86	86	11	95	215
Wisconsin ‡	65	69	27	96	380
Wyoming ‡	92	92	6	94	47
Guam ‡	79	79	8	94	198

National results are based on the national assessment sample, not on aggregated state assessment program samples.
‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for nonpublic school participation rates (see Appendix A).
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Participation Rate Guidelines

In carrying out the 1996 state assessment program, the National Center for Education Statistics (NCES) established participation rate standards that jurisdictions were required to meet in order for their results to be reported. NCES also established additional standards that required the annotation of published results for jurisdictions whose sample participation rates were low enough to raise concerns about their representativeness.

Three states (Nevada, New Hampshire, and New Jersey) failed to meet the initial public school participation rate of 70 percent. For these states, results for eighth-grade public school students are not reported in this or any report of NAEP 1996 science findings. Several other jurisdictions whose results were published received a notation to indicate possible nonresponse bias.

A jurisdiction has its nonpublic school results published in this report and in other reports that include all state-level results if its weighted participation rate for the initial sample of nonpublic schools is greater than or equal to 70 percent AND it meets minimum sample size requirements. Twelve jurisdictions failed to meet one or both of these guidelines at grade 8: Alabama, Connecticut, Delaware, the District of Columbia, Maryland, New Jersey, Oregon, Rhode Island, South Carolina, Utah, Wisconsin, and Wyoming. As with public schools, several other jurisdictions whose nonpublic school results were published received a notation to indicate possible nonresponse bias.

NCES standards require weighted school participation rates before substitution of at least 85 percent to guard against potential bias due to school nonresponse. The NCES standards do not explicitly address the use of substitute schools to replace initially selected schools that declined to participate in the assessment. However, considerable technical consideration has been given to this issue. Even though the characteristics of the substitute schools were matched as closely as possible to the characteristics of the initially selected schools, substitution does not entirely eliminate the possibility of bias because of the nonparticipation of initially selected schools. Thus, for the weighted school participation rates that included substitute schools, the guideline was set at 90 percent. This is expressed in the following guideline:

> *A jurisdiction will receive a notation if its weighted participation rate for the initial sample of schools was below 85 percent* and *the weighted school participation rate after substitution was below 90 percent.*

Seven jurisdictions did not meet this guideline for public schools at grade 8: Arkansas, Iowa, Michigan, Montana, New York, Vermont, and Wisconsin. Fourteen jurisdictions did not meet this guideline for nonpublic schools at grade 8: Arkansas, California, Guam, Kentucky, Louisiana, Massachusetts, Michigan, Minnesota, Nebraska, New Hampshire, New York, North Dakota, Texas, and Vermont.

To help ensure adequate sample representation for each jurisdiction participating in the 1996 state assessment program, NAEP provided substitutes for nonparticipating public and nonpublic schools. (When possible, a substitute school was provided for each initially selected school that declined participation.) For jurisdictions that used substitute schools, the assessment results were based on the student data from all schools participating from both the original sample and the list of substitutes (unless an initial school and its substitute eventually participated, in which case only the data from the initial school were used). For jurisdictions that did not use substitute schools, the participation rates were based on participating schools from the original sample.

The NCES standards specify that attention should be given to the representativeness of the sample coverage. Thus, inadequate representation of an important segment of a jurisdiction's population is of concern, regardless of the overall participation rate. At grade 8, Maryland and South Carolina (for public schools) failed to meet this NCES guideline.

> *A jurisdiction that is not already receiving a notation for problematic overall school or student participation rates will receive a notation if the sampled students within participating schools included a class of students with similar characteristics that had a weighted student response rate of below 80 percent, and from which the nonresponding students together accounted for more than five percent of the jurisdiction's weighted assessable student sample. Student groups from which a jurisdiction needed minimum levels of participation were determined by the age of the students, whether or not the student was classified as a student with a disability (SD) or of limited English proficiency (LEP), and the type of assessment session (monitored or unmonitored). In addition, for public schools, classes of schools were determined by school level of urbanization, minority enrollment, and median household income of the area in which the school is located. For nonpublic schools, classes of schools were determined by type and location of schools.*

This guideline addresses the concern that if nonparticipating schools were concentrated within a particular class of schools, the potential for substantial bias remained, even though the overall level of school participation appeared to be satisfactory. Nonresponse adjustment cells for schools were formed within each jurisdiction, and the schools within each cell were similar in terms of minority enrollment, degree of urbanization, and/or median household income for public schools, and school type and location for nonpublic schools, as appropriate for each jurisdiction. If more than 5 percent (weighted) of the sample schools (after substitution) were nonparticipants from a single adjustment cell, then the potential for nonresponse bias was too great.

In one state (Alaska), the public school student participation rate for grade 8 fell below the NCES-prescribed criterion of 85 percent. No other notations related to student participation rates appear in NAEP 1996 science reports.

Students with Disabilities (SD) and Limited English Proficient (LEP) Students

It is NAEP's intent to assess all selected students. Therefore, every effort is made to ensure that all selected students who are capable of participating in the assessment are assessed. Some students sampled for participation in NAEP can be excluded from the sample according to carefully defined criteria. These criteria are described in Chapter 4 of this report. The results discussed in Chapters 1 through 3 are based on the national and state "reporting samples." For the nation, the reporting sample consisted of schools and students in the S2 sample. For each of the participating jurisdictions, the reporting sample consisted of all schools and students except SD and LEP students from S2 schools. (See Chapter 4 for additional details.) The reporting samples did not allow for the use of accommodations. Sample information for the SD and LEP populations for the reporting samples is presented in Tables A.8a and A.8b.

Table A.8a

NAEP 1996 Reporting Sample SD and LEP Participation Rates: Grade 8, Public Schools

THE NATION'S REPORT CARD naep

	Total Percentage of Students — SD and LEP		Percentage of Students — SD		Percentage of Students — LEP	
	Identified	Excluded	Identified	Excluded	Identified	Excluded
Nation	**14**	**5**	**10**	**4**	**4**	**2**
Alabama	11	7	11	7	0	0
Alaska	16	5	13	4	5	1
Arizona	16	6	9	5	7	2
Arkansas	11	7	10	6	1	1
California	21	9	8	4	14	6
Colorado	12	7	10	5	3	3
Connecticut	15	9	13	8	2	2
Delaware	11	2	10	2	1	1
District of Columbia	12	9	10	7	3	2
Florida	18	10	15	8	4	2
Georgia	11	6	10	5	1	1
Hawaii	11	5	9	4	2	1
Indiana	11	6	11	6	1	0
Iowa	15	6	14	5	1	0
Kentucky	9	4	9	4	0	0
Louisiana	11	6	10	6	0	0
Maine	13	7	13	7	1	0
Maryland	11	5	10	5	2	1
Massachusetts	18	8	15	6	3	2
Michigan	10	5	9	5	1	0
Minnesota	12	4	11	4	2	1
Mississippi	10	6	10	6	0	0
Missouri	13	6	13	6	1	0
Montana	9	3	9	3	0	0
Nebraska	12	4	11	4	1	0
New Mexico	20	9	15	7	7	3
New York	16	9	8	4	8	5
North Carolina	9	5	9	5	1	1
North Dakota	7	2	7	2	0	0
Oregon	12	5	9	4	3	2
Rhode Island	17	7	13	5	5	2
South Carolina	10	6	10	6	0	0
Tennessee	12	4	12	4	1	1
Texas	18	8	11	6	8	3
Utah	9	4	8	4	1	1
Vermont	14	6	13	5	1	1
Virginia	13	7	10	6	4	1
Washington	11	4	8	3	3	1
West Virginia	13	7	13	7	0	0
Wisconsin	12	8	11	7	2	1
Wyoming	11	5	10	4	1	0
DDESS	10	7	8	5	3	3
DoDDS	8	3	6	2	2	1
Guam	11	8	7	5	5	3

National results are based on the national assessment sample, not on aggregated state assessment program samples.

SD = Students with Disabilities (the term previously used was IEP).

LEP = Limited English Proficient Students.

To be excluded, a student was supposed to be classified as SD or as LEP and judged incapable of participating in the assessment.
A student reported as belonging to both SD and LEP classifications is counted once in the overall rate (first column), once in the overall excluded rate (second column), and separately in the remaining columns.

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools

DoDDS: Department of Defense Dependents Schools (Overseas)

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table A.8b

NAEP 1996 Reporting Sample SD and LEP Participation Rates: Grade 8, Nonpublic Schools

THE NATION'S REPORT CARD

	Total Percentage of Students — SD and LEP		Percentage of Students — SD		Percentage of Students — LEP	
	Identified	Excluded	Identified	Excluded	Identified	Excluded
Nation	**3**	**0**	**2**	**0**	**0**	**0**
Arkansas	2	0	2	0	0	0
California	1	0	1	0	0	0
Georgia	0	0	0	0	0	0
Iowa	1	0	1	0	0	0
Kentucky	0	0	0	0	0	0
Louisiana	5	1	5	1	0	0
Massachusetts	5	2	1	0	4	2
Michigan	4	2	3	0	2	2
Minnesota	0	0	0	0	0	0
Missouri	5	0	5	0	0	0
Montana	13	1	1	1	12	0
Nebraska	2	0	1	0	0	0
Nevada	2	2	2	2	0	0
New Hampshire	0	0	0	0	0	0
New Mexico	4	0	4	0	0	0
New York	2	1	2	1	0	0
North Dakota	15	1	6	1	10	1
Texas	4	0	4	0	0	0
Vermont	1	1	0	0	1	1
Washington	1	0	0	0	1	0
Guam	0	0	0	0	0	0

National results are based on the national assessment sample, not on aggregated state assessment program samples.

SD = Students with Disabilities (the term previously used was IEP).

LEP = Limited English Proficient Students.

To be excluded, a student was supposed to be classified as SD or as LEP and judged incapable of participating in the assessment.
A student reported as belonging to both SD and LEP classifications is counted once in the overall rate (first column), once in the overall excluded rate (second column), and separately in the remaining columns.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Scoring

Materials from the 1996 assessment were shipped to National Computer Systems, where trained staff evaluated the responses to the constructed-response questions using scoring rubrics or guides prepared by Educational Testing Service (ETS). Each constructed-response question had a unique scoring guide that defined the criteria used to evaluate students' responses. The extended constructed-response questions were evaluated with four- or five-level guides, while the short constructed-response questions were rated according to two- or three-level guides.

For the national and state science assessments, more than 4.1 million constructed responses were scored. This number includes rescoring to monitor inter-rater reliability. The overall percentages of agreement for the 1996 national reliability samples were 94 percent at grade 4, 94 percent at grade 8, and 93 percent at grade 12.

Data Analysis and IRT Scaling

Subsequent to the professional scoring, all information was transcribed to the NAEP database at ETS. Each processing activity was conducted with rigorous quality control. After the assessment information had been compiled in the database, the data were weighted according to the population structure. The weighting for the national and state samples reflected the probability of selection for each student as a result of the sampling design, adjusted for nonresponse. Through stratification, the weighting assured that the representation of certain subpopulations corresponded to figures from the U.S. Census and the Current Population Survey.[3]

Analyses were then conducted to determine the percentages of students who gave various responses to each cognitive and background question. In determining these percentages for the cognitive questions, a distinction was made between missing responses at the end of a block (i.e., missing responses following the last question the student answered) and missing responses prior to the last observed response. Missing responses before the last observed response were considered intentional omissions. Missing responses at the end of the block were considered "not reached" and treated as if the questions had not been presented to the student. In calculating response percentages for each question, only students classified as having been presented the question were included in the denominator of the statistic.

Item response theory (IRT) was used to estimate average science scale scores for the nation, for various subgroups of interest within the nation, and for the jurisdictions. IRT models the probability of answering a question in a certain way as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale on which performance can be compared across groups — for example, those defined by characteristics such as gender and race/ethnicity.

[3] For additional information about the use of weighting procedures in NAEP, see Johnson, E.G., "Considerations and Techniques for the Analysis of NAEP Data." *Journal of Educational Statistics*, 14(4), pp. 303-334, 1989.

Because of the BIB-spiraling design used by NAEP, students do not receive enough questions about a specific topic to provide reliable information about individual performance. Traditional test scores for individual students, even those based on IRT, would lead to misleading estimates of population characteristics, such as subgroup means and percentages of students at or above a certain scale score level. Consequently, NAEP constructs sets of plausible values designed to represent the distribution of performance in the population. A plausible value for an individual is not a scale score for that individual but may be regarded as a representative value from the distribution of potential scale scores for all students in the population with similar characteristics and identical patterns of item response. Statistics describing performance on the NAEP science scale are based on the plausible values. They estimate values that would have been obtained had individual scale scores been observed — that is, had each student responded to a sufficient number of cognitive questions so his or her individual scores could be precisely estimated.[4]

Three distinct scales were created at each grade to summarize students' abilities in the three defined fields of science: earth, physical, and life. The scales summarize student performance across all three question types in the assessment (multiple-choice, short constructed-response, and extended constructed-response).

The 1996 science assessment was developed using a new framework. Because it was not appropriate to compare results from the 1996 assessment to those of previous NAEP science assessments, no attempt was made to link or align scores on the new assessment to those of previous assessments. Therefore, it was necessary to establish a new scale for reporting. NAEP assessments developed earlier (such as the 1994 reading assessment) were developed with a cross-grade framework, in which the trait being measured is conceptualized as cumulative across the grades of the assessment. Accordingly, a single 0-to-500 scale was established for all three grades in each of these assessments.

In 1993, the National Assessment Governing Board (NAGB) determined that future NAEP assessments should be developed using within-grade frameworks. This removes the constraint that the trait being measured is cumulative. It also means that there is no need for overlap of questions across grades. Consistent with this view, NAGB also declared that scaling be performed within-grade. Any questions which happened to be the same across grades in the assessment were scaled separately for each grade, thus making it possible for common questions to function differently in the separate grades. The NAEP 1994 history and geography assessments were developed and scaled within-grade. After scaling, the scales were aligned so that grade 8 had a higher mean than grade 4, and grade 12 had a higher mean than grade 8. The results were reported on a final 0-to-500 scale that looked similar to those used in reading, in spite of the differences in development and scaling. This definition of the reporting scale was the source of potential confusion and misinterpretation.

[4] For theoretical and empirical justification of the procedures employed, see Mislevy, R.J., "Randomization-Based Inferences about Latent Variables from Complex Samples." *Psychometrika*, 56(2), pp. 177-196, 1988.

For computational details, see National Assessment of Educational Progress, *Focusing the New Design: NAEP 1988 Technical Report* and the *1990 NAEP Technical Report* (Princeton, NJ: Educational Testing Service, 1990).

Therefore, for the NAEP 1996 science assessment — which was also developed and scaled using within-grade procedures — a new reporting metric was adopted. The results are reported on 0-to-300 scales and the means for each of the grades are identical. For each grade, the mean for each field of science was set at 150 and the standard deviation was 35. Constraining the mean and standard deviation of the scales to 150 and 35 also constrained, to some degree, the locations of the percentiles for the total group of students at each grade. However, within-grade comparisons of percentiles across subgroups can still provide valuable comparative information. The reporting metric was developed using data from the national assessment program, and the results for the state assessment program were linked to these scales.

In addition to the plausible values for each scale, a composite of the three fields of science scales was created as a measure of overall science performance. This composite was a weighted average of the plausible values for the three science scales, in which the weights were proportional to the relative importance assigned to each field of science in the assessment framework.

In producing the science scales, three distinct IRT models were used. Multiple-choice questions were scaled using the three-parameter logistic (3PL) model; short constructed-response questions rated as correct or incorrect were scaled using the two-parameter logistic (2PL) model; and short constructed-response questions rated according to a three-level rubric, as well as extended constructed-response questions rated on a four- or five-level rubric, were scaled using a generalized partial-credit (GPC) model.[5] Developed by ETS and first used in 1992, the GPC model permits the scaling of questions scored according to multipoint rating schemes. The model takes full advantage of the information available from each of the student response categories used for these more complex constructed-response questions.

The science scale is composed of three types of questions: multiple-choice questions, constructed-response questions scored dichotomously as correct or incorrect, and constructed-response questions scored according to a partial-credit model. One query about the scale concerns the amount of information contributed by each type of question. Unfortunately, there is no simple answer for the NAEP science assessment, due to the complex procedures used to form the composite science scale.

The information provided by a given question is determined by the IRT model used to scale the question and is a function of its item parameters.[6] Thus, the answer to the query, "How much information do the different types of questions provide?" will differ for each level of science performance. When considering the composite science scale, the answer is even more complicated. The science data are scaled separately by the three fields of science. The composite scale is a weighted combination of these scales. IRT information functions are only strictly comparable when they are derived from the same calibration. Because the composite scale is based on three separate calibrations, there is no direct way to compare the information provided by the questions on the composite scale.

---

[5] Muraki, E., " A Generalized Partial Credit Model: Application of an EM Algorithm." *Applied Psychological Measurement*, 16(2), pp. 159-176, 1992.

[6] Donoghue, J.R., "An Empirical Examination of the IRT Information of Polytomously Scored Reading Items Under the Generalized Partial Credit Model." *Journal of Educational Measurement*, 31(4), pp. 295-311, 1994.

Muraki, E., "Information Functions of the Generalized Partial Credit Model." *Applied Psychological Measurement*, 17(4), pp. 351-363, 1993.

NAEP Reporting Groups

In this report, results are provided for groups of students defined by shared characteristics — region of the country, gender, race/ethnicity, parental education, type of school, participation in Title I programs, and eligibility for the free/reduced-price school lunch program. As described later in this appendix, results are reported for subpopulations only when sufficient numbers of students are assessed and adequate school representation criteria are met. For public school students, the minimum requirement is 62 students in a particular subgroup from at least 5 primary sampling units (PSUs).[7] For nonpublic school students, the minimum requirement is 62 students from at least 6 different schools for the state assessment program or from at least 5 PSUs for the national assessment. The data for all students, regardless of whether their subgroup was reported separately, were included in computing overall results. Definitions of the subpopulations referred to in this report are presented below.

Region

Results are reported for four regions of the nation: Northeast, Southeast, Central, and West. Figure A.2 shows how states are subdivided into these regions. All 50 states and the District of Columbia are listed. Territories and the two Department of Defense Educational Activities jurisdictions are not assigned to any region.

Regional results are based on national assessment samples, not on aggregated state assessment program samples. Thus, the regional results are based on a sample that is different and separate from that used to report the state results.

Figure A.2 Regions of the Country

THE NATION'S REPORT CARD naep

Northeast	Southeast	Central	West
Connecticut	Alabama	Illinois	Alaska
Delaware	Arkansas	Indiana	Arizona
District of Columbia	Florida	Iowa	California
Maine	Georgia	Kansas	Colorado
Maryland	Kentucky	Michigan	Hawaii
Massachusetts	Louisiana	Minnesota	Idaho
New Hampshire	Mississippi	Missouri	Montana
New Jersey	North Carolina	Nebraska	Nevada
New York	South Carolina	North Dakota	New Mexico
Pennsylvania	Tennessee	Ohio	Oklahoma
Rhode Island	Virginia*	South Dakota	Oregon
Vermont	West Virginia	Wisconsin	Texas
Virginia*			Utah
			Washington
			Wyoming

* Note: The part of Virginia that is included in the Washington, DC metropolitan area is included in the Northeast region; the remainder of the state is included in the Southeast region.

[7] For the national assessment, a PSU is a selected geographic region (a county, group of counties, or metropolitan statistical area). For the state assessment program, a PSU is most often a single school.

Gender

Results are reported separately for males and females.

Race/Ethnicity

The race/ethnicity variable is derived from two questions asked of students and, where necessary, school records, and it is used to compare the performance of race/ethnicity subgroups. Two questions from the set of general student background questions were used to determine race/ethnicity:

> If you are Hispanic, what is your Hispanic background?
>
> ❍ I am not Hispanic
> ❍ Mexican, Mexican American, or Chicano
> ❍ Puerto Rican
> ❍ Cuban
> ❍ Other Spanish or Hispanic background

Students who responded to this question by filling in the second, third, fourth, or fifth oval were considered Hispanic. For students who filled in the first oval, did not respond to the question, or provided information that was illegible or could not be classified, responses to the following question were examined to determine their race/ethnicity.

> Which best describes you?
>
> ❍ White (not Hispanic)
> ❍ Black (not Hispanic)
> ❍ Hispanic ("Hispanic" means someone who is Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or other Spanish or Hispanic background)
> ❍ Asian or Pacific Islander ("Asian or Pacific Islander" means someone who is from a Chinese, Japanese, Korean, Filipino, Vietnamese, or other Asian or Pacific Islander background.)
> ❍ American Indian or Alaskan Native ("American Indian or Alaskan Native" means someone who is from one of the American Indian tribes or one of the original people of Alaska.)
> ❍ Other (specify) ______________________________

Students' race/ethnicity was then assigned on the basis of their responses. For students who filled in the sixth oval ("Other") and provided illegible information or information that could not be classified, or who did not respond at all, race/ethnicity was assigned as determined by school records.[8]

Race/ethnicity could not be determined for students who did not respond to either of the demographic questions and whose schools did not provide information about race/ethnicity.

---

[8] The procedure for assigning race/ethnicity was modified for Hawaii. See the forthcoming *Technical Report of the NAEP 1996 State Assessment Program in Science* for details.

Details of how race/ethnicity classifications were derived are presented so that readers can determine how useful the results are for their particular purposes. Also, some students indicated that they were from a Hispanic background (e.g., Puerto Rican or Cuban) and that a racial/ethnic category other than Hispanic best described them. These students were classified as Hispanic based on the rules described above. Furthermore, information from the schools did not always correspond to how students described themselves. Therefore, the racial/ethnic results presented in this report attempt to provide a clear picture based on several sources of information.

Parents' Highest Level of Education

The variable representing the level of parental education is derived from responses to two questions from the set of general student background questions. Students were asked to indicate the extent of their mother's education.

How far in school did your mother go?

- ❍ She did not finish high school.
- ❍ She graduated from high school.
- ❍ She had some education after high school.
- ❍ She graduated from college.
- ❍ I don't know.

Students were asked a similar question about their father's education level.

How far in school did your father go?

- ❍ He did not finish high school.
- ❍ He graduated from high school.
- ❍ He had some education after high school.
- ❍ He graduated from college.
- ❍ I don't know.

The information was combined into one parental education reporting variable through the following process. If a student indicated the extent of education for only one parent, that level was included in the data. If a student indicated the extent of education for both parents, the higher of the two levels was included in the data. If a student did not know the level of education for both parents or did not know the level for one parent and did not respond for the other, the parental education level was classified as "I don't know." (Nationally, 34 percent of fourth graders, 9 percent of eighth graders, and 3 percent of twelfth graders reported that they did not know the education level of either of their parents.) If the student did not respond for either parent, the student was recorded as having provided no response. Approximately 2 percent of the students at each of the three grades provided no response.

Type of School

Results are reported by the type of school that the student attends — public or nonpublic. Nonpublic schools include Catholic and other private schools. Although Bureau of Indian Affairs (BIA) schools and Department of Defense Domestic Dependent Elementary and Secondary Schools (DDESS) are not included in either the public or nonpublic categories, they are included in the overall national results. (A separate sample for DDESS was included as a jurisdiction in the state assessment.)

Students from the overseas Department of Defense Schools (DoDDS) and from the five U.S. Territories (American Samoa, Guam, Northern Marianas, Puerto Rico, and the Virgin Islands) are not included in NAEP national assessment samples. These jurisdictions are eligible, however, to participate in NAEP's state assessment program. Two of these jurisdictions, DoDDS and Guam, as well as DDESS schools, participated as separate jurisdictions, in the 1996 state NAEP program.

Title I Participation

Based on available school records, students were classified either as currently participating in a Title I program or receiving Title I services, or as not receiving such services. The classification applies only to the school year when the assessment was administered (i.e., the 1995–96 school year) and is not based on participation in previous years. If the school did not offer any Title I programs or services, all students in that school were classified as not participating.

Eligibility for the Free/Reduced-Price School Lunch Program

Based on available school records, students were classified as either currently eligible for the free/reduced-price lunch component of the Department of Agriculture's National School Lunch Program or not eligible. The classification applies only to the school year when the assessment was administered (i.e., the 1995–96 school year) and is not based on eligibility in previous years. If school records were not available, the student was classified as "Information not available." If the school did not participate in the program, all students in that school were classified as "Information not available."

Cautions in Interpretations

As described earlier, the NAEP science scale makes it possible to examine relationships between students' performance and various background factors measured by NAEP. However, a relationship that exists between achievement and another variable does not reveal its underlying cause, which may be influenced by a number of other variables. Similarly, the NAEP assessments do not capture the influence of unmeasured variables. The results are most useful when they are considered in combination with other knowledge about the student population and the educational system, such as trends in instruction, changes in the school-age population, and societal demands and expectations.

Guidelines for Analysis and Reporting

This report describes science performance for fourth, eighth, and twelfth graders and compares the results for various groups of students within these populations (e.g., those who have certain demographic characteristics or who responded to a specific background question in a particular way). It also examines the results for individual demographic groups and individual background questions. However, it does not include an analysis of the relationships among combinations of these subpopulations or background questions.

Estimating Variability

Because the statistics presented in this report are estimates of group and subgroup performance based on samples of students rather than the values that could be calculated if every student in the nation answered every question, the degree of uncertainty associated with the estimates should be taken into account. Two components of uncertainty are accounted for in the variability of statistics based on student ability: (1) the uncertainty due to sampling only a relatively small number of students and (2) the uncertainty due to sampling only a relatively small number of cognitive questions. The first component accounts for the variability associated with the estimated percentages of students who had certain background characteristics or who answered a certain cognitive question correctly.

Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate. NAEP uses a jackknife replication procedure to estimate standard errors. The jackknife standard error provides a reasonable measure of uncertainty for any student information that can be observed without error. However, because each student typically responds to only a few questions within any content area, the scale score for any single student would be imprecise. In this case, plausible values technology can be used to describe the performance of groups and subgroups of students, but the underlying imprecision involved in this step adds another component of variability to statistics based on NAEP scale scores.[9] Appendix E provides the standard errors for the results presented in this report.

When the standard error is based on a small number of students or when the group of students is enrolled in a small number of schools, the amount of uncertainty associated with the standard errors may be quite large. Throughout this report, estimates of standard errors subject to a large degree of uncertainty are followed by the "!" symbol. In such cases, the standard errors — and any confidence intervals or significance tests involving these standard errors — should be interpreted cautiously. Additional details concerning procedures for identifying such standard errors are discussed in the forthcoming *NAEP 1996 Technical Report*.

The reader is reminded that, like findings from all surveys, NAEP results are subject to other kinds of error, including the effects of imperfect adjustment for student and school nonresponse and unknowable effects associated with the particular instrumentation and data collection methods. Nonsampling errors can be attributed to a number of sources — inability to

[9] For further details, see Johnson, E.G., & Rust, K.F., "Population Inferences and Variance Estimation for NAEP Data." *Journal of Educational Statistics* 17(2), pp. 175–190, 1992.

obtain complete information about all selected schools in the sample (some students or schools refused to participate, or students participated but answered only certain questions); ambiguous definitions; differences in interpreting questions; inability or unwillingness to give correct information; mistakes in recording, coding, or scoring data; and other errors in data collecting, data processing, and sampling, and in estimating missing data. The extent of nonsampling error is difficult to estimate, and because of their nature, the impact of such errors cannot be reflected in the data-based estimates of uncertainty provided in NAEP reports.

Drawing Inferences from the Results

When the percentages or average scale scores of certain groups are compared, the standard errors should be taken into account, and observed similarities or differences should not be relied on solely. Therefore, the comparisons discussed in this report are based on statistical tests that consider the magnitude of the difference among the averages or percentages and the standard errors of those statistics.

The results from the sample, taking into account the uncertainty associated with all samples, are used to make inferences about the population. Using confidence intervals based on the standard errors provides a way to make inferences about the population averages and percentages in a manner that reflects the uncertainty associated with the sample estimates. An estimated sample average scale score ± 2 standard errors approximates a 95 percent confidence interval for the corresponding population quantity. This statement means that one can conclude at the 95 percent confidence level that the average performance of the entire population of interest (e.g., all fourth-grade students in public schools in a jurisdiction) is within ± 2 standard errors of the sample average.

As an example, suppose that the average science scale score of the students in a particular group was 156 with a standard error of 1.2. A 95 percent confidence interval for the population quantity would be as follows:

$$\begin{array}{l} \text{Average} \pm 2 \text{ standard errors} \\ 156 \pm 2 \times 1.2 \\ 156 \pm 2.4 \\ 153.6,\ 158.4 \end{array}$$

Thus, one can conclude at the 95 percent level of confidence that the average scale score for the entire population of students in that group is between 153.6 and 158.4.

Similar confidence intervals can be constructed for percentages, if the percentages are not extremely large or extremely small. For extreme percentages, confidence intervals constructed in the above manner may not be appropriate, and accurate confidence intervals can be constructed only by using procedures that are quite complicated.

Extreme percentages, defined by both the magnitude of the percentage and the size of the sample from which it was derived, should be interpreted with caution. (The forthcoming *NAEP 1996 Technical Report* contains a more complete discussion of extreme percentages.)

Analyzing Group Differences in Averages and Percentages

The statistical tests determine whether the evidence, based on the data from the groups in the sample, is strong enough to indicate that the averages or percentages are actually different for those groups in the population. If the evidence is strong (i.e., the difference is statistically significant), the report describes the group averages or percentages as being different (e.g., one group performed higher than or lower than another group), regardless of whether the sample averages or percentages appear to be approximately the same. If the evidence is not sufficiently strong (i.e., the difference is not statistically significant), the averages or percentages are described as being not significantly different, regardless of whether the sample averages or percentages appear to be approximately the same or widely discrepant.

Again, the reader is cautioned to rely on the results of the statistical tests rather than on the apparent magnitude of the difference between sample averages or percentages when determining whether the sample differences are likely to represent actual differences among the groups in the population.

To determine whether a real difference exists between the average scale scores (or percentages of a certain attribute) for two groups in the population, one needs to obtain a estimate of the degree of uncertainty associated with the difference between the averages (or percentages) of these groups for the sample. This estimate of the degree of uncertainty, called the standard error of the difference between the groups, is obtained by taking the square of each group's standard error, summing the squared standard errors, and taking the square root of that sum.

$$\text{Standard Error of the Difference} = SE_{A\text{-}B} = \sqrt{SE_A^{\,2} + SE_B^{\,2}}$$

Similar to how the standard error for an individual group average or percentage is used, the standard error of the difference can be used to help determine whether differences among groups in the population are real. The difference between the averages or percentages of the two groups ± two standard errors of the difference represents an approximate 95 percent confidence interval. If the resulting interval includes zero, there is insufficient evidence to claim a real difference between the groups in the population. If the interval does not contain zero, the difference between the groups is statistically significant (different) at the 0.05 level. In this report, differences among groups that involve poorly defined variability estimates (i.e., denoted with "!") and extreme percentages are not discussed.

As an example, to determine whether the average science scale score of Group A is higher that that of Group B, suppose that the sample estimates of the average scale scores and standard errors were as follows:

Group	Average Scale Score	Standard Error
A	118	0.9
B	116	1.1

The difference between the estimates of the average scale scores of Groups A and B is two points (118 - 116). The standard error of this difference is

$$\sqrt{0.9^2 + 1.1^2} = 1.4$$

Thus, an approximate 95 percent confidence interval for this difference is

$$\text{Difference} \pm 2 \text{ standard errors of the difference}$$
$$2 \pm 2 \times 1.4$$
$$2 \pm 2.8$$
$$-0.8, 4.8$$

The value zero is within the confidence interval; therefore, there is insufficient evidence to claim that Group A outperformed Group B.

The procedures described in this section and the certainty ascribed to intervals (e.g., a 95 percent confidence interval) are based on statistical theory that assumes that only one confidence interval or test of statistical significance is being performed. However, in Chapter 2 of this report, many different groups are being compared (i.e., multiple sets of confidence intervals are being analyzed). In sets of confidence intervals, statistical theory indicates that the certainty associated with the entire set of intervals is less than that attributable to each individual comparison from the set. To hold the significance level for the set of comparisons at a particular level (e.g., 0.05), adjustments called multiple comparison procedures must be made to the methods described in the previous section. One such procedure, the Bonferroni method, was used in the analyses described in this report to adjust the confidence intervals for the differences among groups when sets of comparisons were considered.[10] Thus, the confidence intervals for the sets of comparisons in the text are more conservative than those described on the previous pages.

Most of the multiple comparisons in this report pertain to relatively small sets or families of comparisons. For example, for discussions concerning comparisons of parents' level of education, six comparisons were conducted — that is, all pairs of the four parental education levels were compared. In these situations, Bonferroni procedures were appropriate. However, for the cross-state comparisons with a large family of comparisons, the False Discovery Rate (FDR) procedure[11] was used to control the certainty level.

Unlike the Bonferroni procedure, which controls the familywise error rate (i.e., the probability of making even one false rejection in the set of comparisons), the FDR procedure controls the expected proportion of falsely rejected hypotheses. Furthermore, Bonferroni procedures are considered conservative for large families of comparisons.[12] Therefore, the FDR procedure is more suitable for cross-state comparisons. A detailed description of the Bonferroni and FDR procedures appears in the *NAEP 1996 Technical Report* and *Technical Report of the NAEP 1996 State Assessment Program in Science*.

[10] Miller, R.G. *Simultaneous Statistical Inference* (New York: Wiley, 1966).

[11] Benjamini, Y., & Hochberg, Y., "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society*, Series B, 57 (1) pp. 289-300, 1994.

[12] Williams, V. S. L., Jones, L.V., & Tukey, J.W., *Controlling Error in Multiple Comparisons with Special Attention to the National Assessment of Educational Progress* (Research Triangle Park, NC: National Institute of Statistical Sciences, December 1994).

Grade 12 Participation Rates and Motivation

NAEP has been described as a "low-stakes" assessment. That is, students receive no individual scores, and their NAEP performance has no effect on their grades, promotions, or graduation. There has been continued concern that this lack of consequences affects participation rates of students and schools, as well as the motivation of students to perform well on NAEP. Of particular concern has been the performance of twelfth graders, who typically have lower student participation rates than fourth and eighth graders, and who are more likely to omit responses compared to the younger cohorts.

Participation Rates

In NAEP, there has been a consistent pattern of lower participation rates for older students. In the 1994 NAEP assessments, for example, the student participation rates were 93 percent and 91 percent at grades 4 and 8, respectively. At the twelfth grade, however, the participation rate was 81 percent. School participation rates (the percentage of sampled schools that participated in the assessment) have also typically decreased with grade level. In the 1994 assessments, the school participation rate was 86 percent for the fourth grade; 86 percent for the eighth grade; and 79 percent for the twelfth grade.

The effect of participation rates on student performance, however, is unclear. Students may choose not to participate in NAEP for many reasons, such as desire to attend regular classes so as not to miss important instruction, or fear of not doing well on NAEP. Similarly, there are a variety of reasons for which various schools do not participate. The sampling weights and nonresponse adjustments, described earlier in this appendix, provide an approximate statistical adjustment for nonparticipation. However, the effect of some school and student nonparticipation may have an undetermined effect on results.

Motivation

To the extent that students in the NAEP sample are not trying their hardest, NAEP results may underestimate student performance. The concern increases as students get older, and is particularly pronounced for twelfth graders. The students themselves furnish some evidence about their motivation. As part of the background questions, students participating in the 1996 science assessment were asked how important it was for them to do well on the assessment: very important, important, somewhat important, or not very important (see Table A.9). The percentage of students indicating they thought it was either important or very important to do well was 85 percent for fourth graders, 58 percent for eighth graders, and 34 percent for twelfth graders. Motivation to do well decreased at each higher grade assessed.

Table A.9 Students' Report on How Important It Was for Them to Perform Well on the NAEP Science Assessment: Public and Nonpublic Schools Combined

THE NATION'S REPORT CARD naep

	GRADE 4		GRADE 8		GRADE 12	
	Percentage	Average Scale Score	Percentage	Average Scale Score	Percentage	Average Scale Score
Not Very Important	5 (0.3)	137 (2.3)	15 (1.0)	151 (2.2)	29 (0.9)	152 (1.0)
Somewhat Important	9 (0.4)	153 (2.0)	27 (0.6)	154 (1.1)	36 (0.6)	153 (1.2)
Important	26 (0.6)	154 (1.1)	33 (0.7)	151 (1.2)	25 (0.9)	152 (1.9)
Very Important	59 (0.8)	149 (0.8)	25 (1.0)	146 (1.4)	9 (0.5)	139 (2.4)

The standard errors of the estimated percentages and average scale scores appear in parentheses.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Several factors may contribute to this pattern. The NAEP was administered in the late winter, when high-school seniors often have other things on their minds. More recently, the addition to NAEP of more constructed-response questions, which in many instances take longer for the student to answer, may also have had some effect on twelfth graders completing the assessment. As with participation rates, however, the combined effect of these and other factors is unknown.

It is also interesting to note that students who indicated it was very important for them to do well on NAEP did not have the highest average scores. In fact, at grades 8 and 12, students who reported it was not very important to do well had higher average scores than those who reported it was very important to do well. These data further cloud the relationship between motivation and performance on NAEP.

Need for Future Research

More research is needed to delineate the factors that contribute to nonparticipation and lack of motivation. To that end, NCES plans to commission a study of high-school transcripts to learn more about the academic performance of twelfth-grade students who do not participate in the assessment. In addition, NCES is currently investigating how various types of incentives can be effectively used to increase participation in NAEP.

Appendix B

1996 State-Level Results for Selected Subgroups

This appendix includes state-by-state results from the NAEP 1996 state assessment program in science for selected subgroups discussed in Chapter 2. Percentages, average scale scores, and standard errors are presented for gender, race/ethnicity, parental education, type of school (public and nonpublic), Title I participation, and eligibility for the free/reduced-price lunch program. In all the tables in this appendix, DDESS refers to Department of Defense Domestic Dependent Elementary and Secondary Schools and DoDDS refers to overseas Department of Defense Dependents Schools.

Table B.1

Average Science Scale Scores by Gender: Grade 8, Public Schools Only

THE NATION'S REPORT CARD naep

	Male		Female	
	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores
Nation	**51 (1.2)**	**149 (1.1)**	**49 (1.2)**	**148 (1.2)**
Alabama	49 (0.9)	138 (2.0)	51 (0.9)	139 (1.7)
Alaska ‡	50 (1.6)	155 (1.5)	50 (1.6)	150 (1.8)
Arizona	50 (1.1)	147 (1.8)	50 (1.1)	143 (1.7)
Arkansas ‡	50 (1.3)	147 (1.8)	50 (1.3)	142 (1.5)
California	49 (0.9)	140 (2.0)	51 (0.9)	136 (1.9)
Colorado	50 (1.1)	156 (1.2)	50 (1.1)	153 (1.1)
Connecticut	49 (0.9)	156 (1.4)	51 (0.9)	155 (1.5)
Delaware	51 (1.2)	143 (1.4)	49 (1.2)	140 (1.0)
District of Columbia	49 (1.3)	113 (1.2)	51 (1.3)	113 (1.4)
Florida	53 (0.9)	144 (1.6)	47 (0.9)	140 (2.0)
Georgia	50 (1.0)	144 (1.8)	50 (1.0)	139 (1.5)
Hawaii	52 (1.3)	135 (1.0)	48 (1.3)	135 (1.0)
Indiana	50 (1.1)	154 (1.7)	50 (1.1)	152 (1.5)
Iowa ‡	50 (1.1)	159 (1.3)	50 (1.1)	157 (1.4)
Kentucky	50 (1.3)	148 (1.5)	50 (1.3)	147 (1.3)
Louisiana	50 (1.0)	136 (1.9)	50 (1.0)	129 (1.7)
Maine	48 (1.0)	165 (1.2)	52 (1.0)	161 (1.2)
Maryland ‡	51 (1.2)	146 (1.9)	49 (1.2)	145 (1.5)
Massachusetts	52 (1.0)	159 (1.7)	48 (1.0)	154 (1.5)
Michigan ‡	50 (1.2)	156 (1.6)	50 (1.2)	150 (1.7)
Minnesota	50 (1.1)	161 (1.4)	50 (1.1)	157 (1.5)
Mississippi	50 (1.1)	134 (1.8)	50 (1.1)	132 (1.3)
Missouri	51 (1.1)	152 (1.3)	49 (1.1)	150 (1.3)
Montana ‡	49 (1.5)	164 (1.7)	51 (1.5)	160 (1.3)
Nebraska	50 (0.9)	160 (1.2)	50 (0.9)	155 (1.3)
New Mexico	50 (1.0)	143 (1.3)	50 (1.0)	139 (1.1)
New York ‡	50 (1.0)	148 (2.5)	50 (1.0)	143 (1.3)
North Carolina	50 (1.0)	149 (1.5)	50 (1.0)	145 (1.3)
North Dakota	52 (0.9)	163 (0.9)	48 (0.9)	161 (0.9)
Oregon	49 (1.2)	157 (2.0)	51 (1.2)	153 (1.5)
Rhode Island	50 (1.3)	150 (1.1)	50 (1.3)	148 (1.2)
South Carolina ‡	49 (1.1)	141 (1.9)	51 (1.1)	136 (1.5)
Tennessee	52 (1.3)	144 (2.0)	48 (1.3)	142 (2.1)
Texas	50 (1.1)	147 (1.6)	50 (1.1)	143 (2.4)
Utah	48 (1.0)	159 (1.2)	52 (1.0)	154 (0.8)
Vermont ‡	49 (1.4)	158 (1.3)	51 (1.4)	156 (1.1)
Virginia	51 (1.1)	150 (1.7)	49 (1.1)	148 (1.7)
Washington	51 (1.0)	152 (1.6)	49 (1.0)	147 (1.4)
West Virginia	51 (0.9)	148 (1.3)	49 (0.9)	147 (1.1)
Wisconsin ‡	49 (1.2)	161 (1.9)	51 (1.2)	158 (1.7)
Wyoming	52 (1.1)	159 (1.0)	48 (1.1)	156 (0.9)
DDESS	51 (1.9)	157 (1.6)	49 (1.9)	149 (1.6)
DoDDS	49 (1.0)	157 (1.1)	51 (1.0)	154 (0.9)
Guam	50 (1.4)	120 (1.6)	50 (1.4)	120 (1.6)

National results are based on the national assessment sample of public schools, not on aggregated state assessment program samples (see Appendix A).

Scale scores for all grades range from 0 to 300. Standard errors are in parentheses.

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public school participation rates (see Appendix A).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table B.2

Average Science Scale Scores by Race/Ethnicity: Grade 8, Public Schools Only

THE NATION'S REPORT CARD naep

	White		Black		Hispanic		Asian/ Pacific Islander		American Indian	
	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores
Nation	**68 (0.4)**	**159 (1.1)**	**15 (0.3)**	**120 (1.2)**	**12 (0.3)**	**127 (1.8)**	**2 (0.3)**	**150 (3.3)**	**2 (0.3)**	**148 (4.2)**
Alabama	61 (1.9)	151 (1.5)	33 (1.9)	117 (1.8)	4 (0.4)	107 (7.6)	1 (0.3)	*** (***)	2 (0.4)	*** (***)
Alaska ‡	66 (1.6)	162 (1.2)	4 (0.6)	*** (***)	7 (0.8)	137 (4.6)	7 (1.0)	152 (3.8)	16 (1.4)	129 (3.4)
Arizona	57 (1.9)	157 (1.3)	4 (0.6)	124 (3.3)	31 (1.6)	129 (2.1)	2 (0.4)	*** (***)	6 (1.5)	121 (8.6)
Arkansas ‡	73 (1.9)	154 (1.5)	20 (1.7)	116 (2.5)	4 (0.6)	122 (5.8)	1 (0.4)	*** (***)	1 (0.3)	*** (***)
California	38 (2.1)	156 (1.7)	7 (1.0)	121 (3.4)	39 (1.8)	121 (1.9)	13 (1.4)	148 (3.6)	2 (0.3)	*** (***)
Colorado	70 (1.3)	162 (0.8)	5 (0.8)	142 (2.2)	20 (1.2)	135 (2.3)	3 (0.5)	155 (4.8)	3 (0.4)	142 (4.3)
Connecticut	75 (1.4)	165 (1.0)	10 (1.3)	121 (4.4)	11 (0.9)	122 (2.6)	3 (0.4)	163 (3.7)	1 (0.2)	*** (***)
Delaware	64 (1.2)	152 (0.8)	26 (1.0)	122 (1.8)	7 (0.7)	116 (4.1)	2 (0.3)	*** (***)	2 (0.3)	*** (***)
District of Columbia	3 (0.3)	*** (***)	83 (0.9)	112 (0.9)	11 (0.8)	98 (3.3)	1 (0.4)	*** (***)	1 (0.2)	*** (***)
Florida	55 (2.1)	155 (1.5)	20 (2.0)	119 (2.7)	22 (2.0)	129 (2.2)	2 (0.4)	*** (***)	1 (0.2)	*** (***)
Georgia	56 (2.3)	155 (1.2)	36 (2.4)	122 (1.4)	5 (0.4)	128 (4.2)	2 (0.4)	*** (***)	1 (0.3)	*** (***)
Hawaii	17 (0.7)	146 (1.8)	3 (0.4)	128 (4.4)	22 (0.8)	121 (1.8)	54 (1.3)	138 (1.1)	2 (0.3)	*** (***)
Indiana	81 (1.8)	158 (1.3)	11 (1.4)	125 (3.3)	5 (0.7)	139 (2.1)	1 (0.2)	*** (***)	2 (0.4)	*** (***)
Iowa ‡	91 (1.0)	160 (1.1)	3 (0.6)	131 (3.6)	3 (0.5)	140 (4.6)	2 (0.3)	*** (***)	1 (0.2)	*** (***)
Kentucky	86 (0.9)	151 (1.1)	9 (0.8)	127 (2.7)	3 (0.4)	113 (6.2)	1 (0.2)	*** (***)	1 (0.2)	*** (***)
Louisiana	55 (1.8)	148 (1.3)	37 (1.7)	113 (2.1)	6 (0.6)	104 (5.7)	1 (0.3)	*** (***)	1 (0.3)	*** (***)
Maine	92 (0.7)	164 (0.9)	1 (0.2)	*** (***)	3 (0.5)	141 (4.6)	1 (0.3)	*** (***)	2 (0.3)	*** (***)
Maryland ‡	56 (2.0)	160 (1.4)	32 (2.1)	124 (1.4)	6 (0.6)	121 (4.1)	4 (0.6)	161 (3.6)	2 (0.3)	*** (***)
Massachusetts	81 (1.7)	163 (1.2)	6 (1.0)	126 (3.3)	8 (0.7)	126 (3.9)	4 (0.8)	152 (7.3)	1 (0.2)	*** (***)
Michigan ‡	76 (2.0)	161 (1.4)	15 (1.9)	122 (2.4)	4 (0.4)	134 (4.9)	2 (0.5)	*** (***)	2 (0.3)	*** (***)
Minnesota	86 (1.9)	162 (1.2)	4 (0.8)	130 (4.4)	4 (0.6)	134 (5.3)	4 (0.9)	152 (9.7)	2 (0.5)	*** (***)
Mississippi	50 (2.1)	149 (1.2)	44 (1.9)	119 (1.4)	6 (0.6)	105 (3.8)	0 (0.1)	*** (***)	1 (0.2)	*** (***)
Missouri	78 (1.5)	158 (1.0)	13 (1.3)	120 (2.8)	5 (0.6)	130 (5.0)	1 (0.3)	*** (***)	2 (0.4)	*** (***)
Montana ‡	83 (1.9)	166 (0.9)	1 (0.1)	*** (***)	5 (0.5)	147 (2.7)	1 (0.2)	*** (***)	10 (1.7)	139 (2.7)
Nebraska	85 (1.2)	161 (0.9)	5 (0.6)	130 (3.1)	7 (0.9)	134 (3.1)	1 (0.2)	*** (***)	2 (0.3)	*** (***)
New Mexico	38 (1.5)	159 (1.0)	3 (0.4)	*** (***)	51 (1.5)	130 (1.1)	1 (0.2)	*** (***)	8 (0.6)	126 (2.4)
New York ‡	60 (2.6)	161 (1.4)	17 (2.0)	120 (1.9)	16 (1.2)	116 (2.7)	5 (0.9)	155 (5.4)	2 (0.5)	*** (***)
North Carolina	65 (2.0)	157 (1.1)	27 (1.3)	126 (1.4)	4 (0.5)	123 (3.6)	1 (0.3)	*** (***)	3 (1.4)	136 (4.1)
North Dakota	92 (0.8)	164 (0.8)	1 (0.2)	*** (***)	4 (0.4)	137 (4.5)	1 (0.2)	*** (***)	3 (0.7)	137 (6.9)
Oregon	82 (1.5)	158 (1.4)	2 (0.5)	*** (***)	8 (1.0)	133 (3.7)	4 (0.5)	157 (3.3)	4 (0.8)	142 (7.9)
Rhode Island	77 (0.8)	155 (0.9)	5 (0.5)	130 (2.8)	12 (0.5)	118 (1.8)	4 (0.4)	142 (3.1)	1 (0.2)	*** (***)
South Carolina ‡	51 (1.9)	153 (1.6)	40 (1.9)	122 (1.6)	6 (0.6)	122 (4.1)	1 (0.3)	*** (***)	2 (0.3)	*** (***)
Tennessee	77 (1.5)	151 (1.7)	17 (1.5)	117 (3.1)	3 (0.5)	104 (6.2)	1 (0.2)	*** (***)	1 (0.3)	*** (***)
Texas	48 (1.9)	161 (1.2)	12 (1.3)	127 (2.4)	36 (2.1)	129 (2.7)	3 (0.5)	157 (3.6)	1 (0.2)	*** (***)
Utah	87 (1.0)	159 (0.7)	1 (0.2)	*** (***)	8 (0.7)	133 (2.9)	3 (0.4)	143 (3.2)	1 (0.3)	*** (***)
Vermont ‡	90 (0.9)	159 (0.9)	1 (0.3)	*** (***)	4 (0.5)	136 (3.4)	1 (0.3)	*** (***)	3 (0.5)	*** (***)
Virginia	64 (2.0)	158 (1.4)	24 (1.9)	126 (2.3)	5 (0.6)	132 (4.2)	5 (0.6)	165 (3.2)	1 (0.3)	*** (***)
Washington	74 (1.9)	156 (1.1)	4 (0.7)	127 (4.2)	10 (1.1)	125 (3.5)	7 (0.9)	149 (3.3)	4 (0.6)	130 (4.3)
West Virginia	90 (0.7)	149 (0.9)	4 (0.5)	127 (3.2)	3 (0.3)	122 (4.3)	1 (0.2)	*** (***)	2 (0.3)	*** (***)
Wisconsin ‡	83 (1.5)	165 (1.1)	6 (1.1)	115 (5.3)	6 (0.7)	141 (4.6)	2 (0.4)	*** (***)	2 (0.5)	*** (***)
Wyoming	84 (0.8)	161 (0.6)	1 (0.2)	*** (***)	11 (0.6)	140 (1.9)	1 (0.2)	*** (***)	4 (0.4)	138 (2.5)
DDESS	47 (1.7)	162 (1.7)	22 (1.5)	137 (2.5)	24 (1.3)	149 (2.4)	3 (0.9)	*** (***)	2 (0.5)	*** (***)
DoDDS	45 (0.9)	164 (1.2)	19 (0.8)	140 (1.2)	17 (0.8)	146 (1.6)	14 (0.7)	156 (1.4)	2 (0.3)	*** (***)
Guam	8 (0.9)	138 (4.6)	3 (0.6)	*** (***)	19 (1.3)	106 (2.9)	69 (1.6)	122 (1.4)	0 (0.2)	*** (***)

National results are based on the national assessment sample of public schools, not on aggregated state assessment program samples (see Appendix A).

Scale scores for all grades range from 0 to 300. Standard errors are in parentheses.

*** Sample size insufficient to permit reliable estimates.

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public school participation rates (see Appendix A).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table B.3

Average Science Scale Scores by Parents' Highest Level of Education: Grade 8, Public Schools Only

THE NATION'S REPORT CARD naep

	Did Not Finish High School		Graduated From High School		Some Education After High School		Graduated From College		I Don't Know	
	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores
Nation	**7 (0.5)**	**131 (2.0)**	**21 (1.0)**	**140 (1.5)**	**20 (0.7)**	**155 (1.2)**	**42 (1.3)**	**157 (1.3)**	**10 (0.6)**	**133 (2.6)**
Alabama	8 (0.7)	130 (3.0)	25 (1.0)	129 (2.0)	18 (0.9)	145 (1.7)	42 (1.9)	147 (2.3)	8 (0.7)	122 (2.7)
Alaska ‡	4 (0.7)	*** (***)	15 (1.2)	141 (3.1)	24 (1.3)	155 (1.5)	46 (1.4)	163 (1.3)	11 (1.0)	132 (4.1)
Arizona	9 (0.8)	121 (3.1)	17 (1.2)	136 (2.1)	22 (1.0)	151 (1.7)	40 (1.8)	158 (1.4)	12 (0.9)	128 (2.6)
Arkansas ‡	9 (1.0)	129 (3.3)	25 (1.2)	136 (1.9)	24 (1.1)	150 (1.9)	33 (1.7)	154 (2.0)	10 (0.7)	133 (4.2)
California	10 (0.9)	118 (2.7)	17 (1.1)	129 (2.5)	16 (0.9)	144 (2.0)	40 (1.9)	153 (2.0)	17 (1.1)	118 (2.8)
Colorado	5 (0.5)	133 (3.9)	16 (0.9)	142 (1.9)	20 (1.0)	157 (1.6)	51 (1.5)	163 (0.9)	8 (0.6)	136 (2.3)
Connecticut	5 (0.5)	129 (3.8)	18 (1.0)	140 (2.3)	18 (0.8)	155 (1.7)	52 (1.4)	167 (1.2)	9 (0.5)	132 (2.9)
Delaware	5 (0.6)	121 (4.5)	26 (1.1)	135 (1.8)	20 (1.0)	146 (1.4)	41 (1.2)	151 (1.4)	9 (0.8)	122 (4.9)
District of Columbia	6 (0.6)	106 (3.1)	27 (1.1)	107 (1.8)	16 (0.7)	120 (2.5)	37 (1.2)	121 (1.6)	14 (1.0)	100 (2.2)
Florida	7 (0.7)	127 (3.4)	19 (1.1)	132 (2.3)	21 (1.1)	148 (1.5)	42 (1.4)	150 (2.0)	11 (0.8)	127 (2.5)
Georgia	8 (0.7)	127 (2.4)	24 (1.3)	129 (2.1)	19 (1.0)	145 (1.6)	43 (2.0)	153 (2.2)	7 (0.5)	128 (2.8)
Hawaii	4 (0.4)	119 (5.3)	24 (1.0)	120 (2.3)	18 (0.9)	139 (1.9)	39 (0.9)	147 (1.1)	15 (1.0)	129 (1.9)
Indiana	5 (0.5)	139 (2.9)	27 (1.1)	144 (1.9)	21 (1.2)	156 (1.7)	41 (1.9)	162 (1.9)	6 (0.6)	135 (3.8)
Iowa ‡	4 (0.5)	141 (3.4)	20 (1.0)	150 (1.5)	20 (0.8)	160 (1.7)	48 (1.5)	165 (1.2)	8 (0.7)	141 (3.5)
Kentucky	11 (0.6)	130 (2.1)	27 (1.1)	143 (1.5)	23 (1.0)	151 (1.6)	32 (1.5)	158 (1.8)	8 (0.5)	134 (2.6)
Louisiana	9 (0.6)	123 (3.1)	29 (1.1)	128 (1.9)	20 (0.8)	141 (2.1)	35 (1.5)	136 (2.3)	8 (0.6)	124 (3.0)
Maine	4 (0.6)	141 (2.9)	20 (1.1)	153 (1.5)	22 (1.1)	164 (1.7)	48 (1.5)	171 (1.1)	6 (0.6)	148 (2.6)
Maryland ‡	5 (0.5)	126 (3.6)	20 (1.2)	136 (2.0)	18 (0.8)	147 (2.0)	48 (1.7)	153 (2.0)	8 (0.6)	134 (2.6)
Massachusetts	4 (0.5)	134 (4.7)	17 (1.0)	145 (2.4)	16 (0.9)	156 (2.0)	56 (1.8)	166 (1.3)	8 (0.7)	134 (2.8)
Michigan ‡	4 (0.5)	137 (5.3)	20 (1.0)	144 (2.0)	21 (1.1)	156 (1.6)	46 (1.7)	161 (1.7)	9 (0.7)	135 (3.2)
Minnesota	3 (0.3)	137 (4.5)	18 (1.1)	151 (1.8)	22 (1.1)	161 (1.7)	50 (1.8)	165 (1.4)	7 (0.6)	142 (3.9)
Mississippi	8 (0.6)	125 (2.5)	24 (0.9)	126 (1.9)	16 (0.7)	142 (1.8)	42 (1.3)	138 (1.9)	10 (0.5)	119 (2.6)
Missouri	7 (0.6)	136 (2.8)	25 (1.1)	144 (1.6)	21 (1.0)	156 (1.4)	39 (1.5)	159 (1.3)	7 (0.6)	135 (3.2)
Montana ‡	5 (0.5)	139 (3.1)	19 (1.4)	155 (2.2)	22 (0.8)	164 (1.5)	48 (1.4)	168 (1.3)	6 (0.6)	147 (3.6)
Nebraska	4 (0.5)	133 (2.8)	20 (0.9)	148 (1.8)	18 (0.8)	161 (1.5)	50 (1.1)	165 (1.2)	8 (0.6)	136 (2.8)
New Mexico	9 (0.7)	119 (2.4)	21 (0.9)	131 (1.8)	20 (0.6)	147 (1.5)	39 (1.2)	154 (1.2)	10 (0.7)	125 (2.3)
New York ‡	6 (0.7)	123 (5.4)	16 (0.9)	138 (3.6)	19 (1.1)	147 (2.0)	49 (1.4)	157 (1.7)	11 (0.8)	124 (2.4)
North Carolina	7 (0.5)	126 (2.6)	22 (1.1)	134 (1.7)	21 (0.8)	150 (1.7)	42 (1.5)	158 (1.4)	8 (0.7)	133 (2.3)
North Dakota	3 (0.4)	148 (3.7)	16 (0.8)	157 (1.9)	18 (0.8)	160 (1.6)	57 (1.0)	167 (0.9)	6 (0.6)	146 (3.5)
Oregon	6 (0.7)	137 (3.1)	16 (1.0)	143 (2.0)	22 (1.0)	157 (1.5)	47 (1.6)	164 (1.7)	9 (0.9)	135 (4.1)
Rhode Island	8 (0.6)	123 (2.7)	17 (0.9)	141 (1.9)	18 (0.9)	154 (1.8)	45 (1.3)	160 (1.0)	12 (0.8)	130 (2.6)
South Carolina ‡	7 (0.7)	125 (3.7)	26 (1.3)	127 (1.8)	17 (1.0)	145 (2.1)	41 (1.5)	148 (2.1)	9 (0.7)	127 (3.0)
Tennessee	10 (0.8)	127 (2.4)	28 (1.4)	135 (2.2)	21 (0.9)	149 (2.2)	36 (1.9)	154 (2.2)	6 (0.6)	129 (3.6)
Texas	13 (0.9)	128 (2.0)	19 (1.0)	137 (2.4)	19 (1.0)	152 (1.8)	39 (1.5)	157 (1.5)	10 (0.8)	125 (3.3)
Utah	2 (0.4)	129 (5.9)	15 (0.9)	147 (1.5)	20 (0.8)	156 (1.5)	54 (1.0)	162 (0.8)	8 (0.4)	138 (1.9)
Vermont ‡	5 (0.6)	132 (4.3)	23 (1.3)	146 (1.5)	17 (0.9)	157 (1.8)	50 (1.4)	167 (1.1)	6 (0.6)	143 (3.4)
Virginia	7 (0.7)	127 (2.8)	20 (1.1)	136 (2.1)	18 (0.7)	152 (1.9)	47 (1.6)	161 (1.9)	8 (0.8)	137 (3.5)
Washington	6 (0.8)	128 (4.2)	15 (0.9)	141 (2.3)	21 (0.7)	154 (1.7)	48 (1.6)	158 (1.4)	11 (0.9)	133 (3.2)
West Virginia	9 (0.6)	130 (2.3)	29 (1.1)	142 (1.2)	21 (0.8)	152 (1.3)	33 (1.0)	156 (1.3)	7 (0.5)	134 (2.8)
Wisconsin ‡	4 (0.4)	140 (4.3)	23 (1.3)	155 (2.3)	24 (1.0)	161 (1.8)	40 (1.7)	169 (1.6)	8 (0.7)	138 (3.6)
Wyoming	5 (0.4)	139 (2.7)	18 (0.7)	150 (1.3)	23 (1.0)	159 (1.3)	46 (1.0)	165 (0.9)	8 (0.6)	143 (3.1)
DDESS	3 (0.8)	*** (***)	15 (1.5)	142 (3.0)	24 (1.5)	153 (2.0)	51 (2.3)	158 (1.7)	7 (1.2)	*** (***)
DoDDS	1 (0.2)	*** (***)	12 (0.8)	144 (1.9)	23 (0.8)	159 (1.3)	53 (1.0)	158 (1.0)	10 (0.7)	146 (2.0)
Guam	7 (0.9)	106 (3.6)	28 (1.6)	113 (2.0)	17 (0.9)	130 (2.4)	34 (1.5)	128 (2.1)	14 (1.3)	110 (3.3)

National results are based on the national assessment sample of public schools, not on aggregated state assessment program samples (see Appendix A).

Scale scores for all grades range from 0 to 300. Standard errors are in parentheses.

*** Sample size insufficient to permit reliable estimates.

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public school participation rates (see Appendix A).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table B.4

Average Science Scale Scores by Type of School: Grade 8

THE NATION'S REPORT CARD naep

	Public		Nonpublic	
	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores
Nation	**89 (1.4)**	**148 (0.9)**	**11 (1.4)**	**162 (2.5)**
Arkansas ‡	96 (1.1)	144 (1.3)	4 (1.1)	167 (4.4)
California ‡	92 (1.0)	138 (1.7)	8 (1.0)	161 (4.3)
Georgia	94 (1.8)	142 (1.4)	6 (1.8)	166 (5.2)
Iowa ‡	92 (1.4)	158 (1.2)	9 (1.4)	167 (3.2)
Kentucky ‡	91 (1.3)	147 (1.2)	9 (1.3)	159 (3.7)
Louisiana ‡	83 (1.8)	132 (1.6)	17 (1.8)	156 (3.2)
Massachusetts ‡	85 (1.4)	157 (1.4)	15 (1.4)	161 (3.3)
Michigan ‡	88 (1.0)	153 (1.4)	12 (1.0)	158 (4.0)
Minnesota ‡	92 (1.0)	159 (1.3)	8 (1.0)	166 (2.4)
Missouri	88 (1.7)	151 (1.2)	12 (1.7)	167 (4.2)
Montana ‡	95 (1.1)	162 (1.2)	5 (1.1)	158 (8.6)
Nebraska ‡	89 (1.4)	157 (1.0)	11 (1.4)	165 (2.5)
New Mexico	93 (1.4)	141 (1.0)	7 (1.4)	164 (6.6)
New York ‡	84 (1.5)	146 (1.6)	16 (1.5)	149 (4.7)
North Dakota ‡	94 (1.4)	162 (0.8)	7 (1.4)	168 (4.5)
Texas ‡	95 (0.8)	145 (1.8)	5 (0.8)	176 (9.2)
Vermont ‡	95 (0.8)	157 (1.0)	5 (0.8)	168 (4.6)
Washington	93 (1.1)	150 (1.3)	7 (1.1)	165 (6.0)
Guam ‡	80 (0.8)	120 (1.1)	20 (0.8)	147 (1.8)

Results are presented for jurisdictions with reportable public and nonpublic school results (see Appendix A).

National results are based on the national assessment samples, not on aggregated state assessment program samples.

Scale scores for all grades range from 0 to 300. Standard errors are in parentheses.

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public and/or nonpublic school participation rates (see Appendix A).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table B.5

Average Science Scale Scores by Title I Participation: Grade 8, Public Schools Only

THE NATION'S REPORT CARD naep

	Participated		Did Not Participate	
	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores
Nation	**13 (2.3)**	**127 (4.9)**	**87 (2.3)**	**152 (1.2)**
Alabama	16 (2.5)	117 (3.3)	84 (2.5)	143 (1.7)
Alaska ‡	4 (1.8)	(***) (***)	96 (1.8)	155 (1.2)
Arizona	16 (2.5)	125 (4.2)	84 (2.5)	149 (1.7)
Arkansas ‡	17 (2.3)	124 (3.3)	83 (2.3)	148 (1.6)
California	26 (3.2)	112 (2.6)	74 (3.2)	147 (1.6)
Colorado	2 (0.6)	(***) (***)	98 (0.6)	155 (0.8)
Connecticut	4 (1.1)	127 (3.6)	96 (1.1)	156 (1.4)
Delaware	0 (0.1)	(***) (***)	100 (0.1)	142 (0.8)
District of Columbia	15 (0.7)	101 (2.9)	85 (0.7)	115 (0.8)
Florida	9 (2.8)	115 (5.5)	91 (2.8)	145 (1.4)
Georgia	11 (1.3)	115 (4.5)	89 (1.3)	145 (1.6)
Hawaii	8 (0.5)	111 (1.8)	92 (0.5)	137 (0.8)
Indiana	2 (0.7)	(***) (***)	98 (0.7)	154 (1.3)
Iowa ‡	1 (0.4)	(***) (***)	99 (0.4)	158 (1.2)
Kentucky	20 (2.3)	132 (2.1)	80 (2.3)	151 (1.3)
Louisiana	14 (2.4)	119 (4.0)	86 (2.4)	135 (1.7)
Maine	4 (0.8)	143 (2.7)	96 (0.8)	164 (1.0)
Maryland ‡	2 (0.8)	(***) (***)	98 (0.8)	146 (1.4)
Massachusetts	11 (1.8)	125 (3.8)	89 (1.8)	161 (1.5)
Michigan ‡	15 (1.9)	129 (4.7)	85 (1.9)	157 (1.6)
Minnesota	3 (0.7)	131 (5.7)	97 (0.7)	159 (1.3)
Mississippi	33 (3.0)	120 (2.1)	67 (3.0)	139 (1.6)
Missouri	8 (1.4)	116 (5.2)	92 (1.4)	154 (1.0)
Montana ‡	9 (1.1)	137 (2.7)	91 (1.1)	164 (1.3)
Nebraska	2 (0.7)	(***) (***)	98 (0.7)	158 (1.0)
New Mexico	15 (1.7)	117 (2.2)	85 (1.7)	145 (0.9)
New York ‡	16 (2.6)	115 (3.2)	84 (2.6)	152 (1.8)
North Carolina	6 (2.0)	123 (4.1)	94 (2.0)	148 (1.2)
North Dakota	6 (0.8)	129 (3.1)	94 (0.8)	164 (0.7)
Oregon	4 (0.9)	128 (4.3)	96 (0.9)	156 (1.5)
Rhode Island	9 (0.5)	115 (2.0)	91 (0.5)	152 (0.8)
South Carolina ‡	8 (2.7)	123 (3.9)	92 (2.7)	140 (1.5)
Tennessee	5 (2.0)	112 (7.0)	95 (2.0)	145 (1.9)
Texas	22 (2.8)	124 (2.1)	78 (2.8)	151 (2.2)
Utah	3 (0.6)	121 (4.5)	97 (0.6)	157 (0.8)
Vermont ‡	6 (0.8)	131 (3.1)	94 (0.8)	159 (0.9)
Virginia	1 (0.4)	(***) (***)	99 (0.4)	150 (1.6)
Washington	8 (1.4)	128 (5.4)	92 (1.4)	151 (1.2)
West Virginia	8 (1.6)	125 (3.2)	92 (1.6)	149 (0.9)
Wisconsin ‡	9 (2.3)	120 (5.8)	91 (2.3)	164 (1.1)
Wyoming	4 (0.4)	135 (2.0)	96 (0.4)	158 (0.7)
DDESS	0 (—) !	(***) (***)	100 (—) !	153 (1.1)
DoDDS	2 (0.3)	(***) (***)	98 (0.3)	155 (0.7)
Guam	0 (—) !	(***) (***)	100 (—) !	120 (1.1)

National results are based on the national assessment sample of public schools, not on aggregated state assessment program samples (see Appendix A).

Scale scores for all grades range from 0 to 300. Standard errors are in parentheses.

*** Sample size insufficient to permit reliable estimates.

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public school participation rates (see Appendix A).

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

— Standard error estimates cannot be accurately determined.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table B.6 — Average Science Scale Scores by Free/Reduced-Price Lunch Eligibility: Grade 8, Public Schools Only

THE NATION'S REPORT CARD naep

	Eligible		Not Eligible		Information Not Available	
	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores	Percentage of Students	Average Scale Scores
Nation	**29 (1.6)**	**133 (1.7)**	**51 (3.6)**	**155 (1.3)**	**20 (4.4)**	**154 (3.6)**
Alabama	39 (1.9)	121 (1.9)	58 (2.0)	150 (1.7)	3 (1.2)	151 (9.3)
Alaska ‡	20 (1.6)	131 (3.7)	34 (1.4)	157 (1.7)	47 (1.8)	159 (1.8)
Arizona	28 (2.5)	127 (2.8)	52 (3.7)	155 (1.7)	20 (3.9)	144 (2.0)
Arkansas ‡	33 (1.8)	128 (1.7)	60 (2.8)	152 (1.3)	6 (3.1)	155 (9.0)
California	36 (2.6)	120 (2.0)	47 (3.0)	152 (2.0)	17 (3.1)	137 (4.0)
Colorado	24 (1.8)	137 (1.9)	65 (2.5)	160 (1.0)	11 (2.5)	157 (3.1)
Connecticut	21 (1.5)	127 (3.3)	74 (2.1)	163 (1.1)	5 (1.6)	154 (10.9)
Delaware	22 (1.1)	119 (2.3)	56 (1.0)	152 (0.9)	22 (0.4)	137 (1.4)
District of Columbia	55 (1.2)	107 (1.2)	30 (0.9)	124 (1.8)	15 (0.8)	114 (2.3)
Florida	39 (1.9)	127 (1.9)	53 (2.9)	154 (1.5)	8 (2.6)	138 (5.0)
Georgia	32 (2.3)	124 (1.6)	54 (2.7)	151 (1.6)	14 (3.5)	146 (5.7)
Hawaii	29 (1.0)	125 (1.7)	66 (1.0)	141 (0.9)	5 (0.3)	115 (2.1)
Indiana	21 (1.5)	136 (2.3)	79 (1.6)	158 (1.3)	1 (0.3)	(***) (***)
Iowa ‡	21 (1.3)	144 (1.9)	73 (2.4)	162 (1.2)	6 (2.2)	155 (2.7)
Kentucky	34 (2.1)	135 (1.6)	59 (2.3)	155 (1.3)	7 (2.5)	142 (3.3)
Louisiana	48 (2.1)	121 (1.9)	45 (1.9)	145 (1.5)	7 (2.0)	128 (7.5)
Maine	24 (1.3)	152 (1.7)	71 (1.8)	167 (1.0)	5 (1.8)	164 (3.4)
Maryland ‡	26 (1.9)	122 (2.1)	69 (2.6)	154 (1.7)	5 (2.2)	143 (6.6)
Massachusetts	18 (1.5)	133 (1.8)	73 (3.0)	164 (1.2)	9 (2.8)	149 (6.8)
Michigan ‡	19 (1.8)	139 (1.9)	66 (3.8)	159 (1.5)	14 (4.2)	144 (8.3)
Minnesota	20 (1.5)	145 (2.4)	64 (3.1)	162 (1.1)	16 (3.1)	162 (5.0)
Mississippi	52 (1.9)	121 (1.5)	42 (2.0)	148 (1.5)	6 (2.5)	134 (5.6)
Missouri	27 (1.6)	138 (1.9)	65 (2.6)	157 (1.0)	8 (2.7)	144 (8.0)
Montana ‡	25 (1.8)	150 (2.0)	60 (2.8)	166 (1.2)	16 (2.8)	165 (1.9)
Nebraska	27 (1.6)	144 (1.6)	69 (1.8)	162 (0.9)	5 (1.0)	161 (5.3)
New Mexico	41 (1.5)	130 (1.5)	43 (1.9)	151 (1.1)	16 (1.5)	143 (2.4)
New York ‡	37 (2.3)	124 (1.9)	54 (2.8)	159 (1.8)	9 (2.6)	153 (7.1)
North Carolina	31 (1.8)	128 (1.4)	62 (2.1)	156 (1.2)	8 (2.4)	144 (3.4)
North Dakota	20 (1.1)	157 (1.5)	70 (1.7)	165 (0.7)	10 (1.6)	155 (3.6)
Oregon	23 (1.5)	145 (2.0)	64 (3.0)	159 (1.5)	13 (3.0)	151 (5.6)
Rhode Island	25 (0.8)	131 (1.4)	71 (0.7)	157 (0.9)	4 (0.2)	125 (3.1)
South Carolina ‡	45 (2.2)	126 (1.8)	54 (2.0)	149 (1.4)	1 (—) !	(***) (***)
Tennessee	28 (2.3)	125 (2.4)	64 (2.5)	151 (2.0)	8 (2.3)	144 (5.3)
Texas	37 (2.2)	130 (1.7)	56 (2.6)	157 (1.3)	6 (2.0)	127 (15.1)
Utah	20 (1.3)	149 (1.7)	69 (1.7)	158 (0.9)	11 (1.6)	157 (2.0)
Vermont ‡	20 (1.1)	146 (2.1)	73 (1.7)	160 (0.9)	7 (1.8)	157 (2.9)
Virginia	21 (1.7)	125 (2.2)	67 (2.8)	157 (1.6)	12 (3.0)	150 (4.5)
Washington	24 (1.6)	135 (2.1)	73 (2.1)	154 (1.2)	3 (1.5)	155 (3.7)
West Virginia	35 (1.5)	138 (1.3)	61 (2.0)	152 (1.0)	4 (1.9)	151 (4.8)
Wisconsin ‡	21 (2.0)	140 (3.5)	65 (4.0)	166 (1.2)	14 (4.0)	161 (3.8)
Wyoming	20 (0.8)	148 (1.2)	75 (0.8)	160 (0.8)	5 (0.4)	155 (4.8)
DDESS	24 (1.9)	148 (2.0)	43 (1.9)	158 (1.8)	33 (0.8)	150 (2.1)
DoDDS	7 (0.5)	146 (2.4)	49 (0.7)	156 (0.9)	44 (0.4)	156 (1.1)
Guam	18 (1.2)	101 (2.2)	81 (1.3)	125 (1.1)	1 (0.2)	(***) (***)

National results are based on the national assessment sample of public schools, not on aggregated state assessment program samples (see Appendix A).

Scale scores for all grades range from 0 to 300. Standard errors are in parentheses.

*** Sample size insufficient to permit reliable estimates.

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public school participation rates (see Appendix A).

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistics does not match statistical test assumptions (See Appendix A).

— Standard error estimates cannot be accurately determined.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Appendix C

State-Level Contextual Variables

To help place results from the NAEP 1996 state assessment program into context, this appendix presents selected state-level data from sources other than NAEP. The information presented is taken from the *Digest of Education Statistics 1996*.

Table C.1	School System Characteristics from Non-NAEP Sources

	Estimated Total and School-Age Resident Population: 1995 (Estimates as of July 1)[1]		Enrollment in Public Elementary and Secondary Schools: Fall 1994[2]		
	Total, All Ages (in thousands)	5- to 17-year-olds (in thousands)	Total	Kindergarten through Grade 8	Grades 9 to 12
United States	262,755	49,149	44,108,775	31,894,333	12,214,442
Alabama	4,253	779	736,472	535,187	201,285
Alaska	604	136	127,057	93,719	33,338
Arizona	4,218	837	737,424	542,904	194,520
Arkansas	2,484	477	447,565	319,282	128,283
California	31,589	5,984	5,407,043	3,955,434	1,451,609
Colorado	3,747	712	640,521	469,755	170,766
Connecticut	3,275	570	506,824	375,638	131,186
Delaware	717	127	106,813	76,819	29,994
District of Columbia	554	75	80,450	62,126	18,324
Florida	14,166	2,403	2,108,968	1,567,328	541,640
Georgia	7,201	1,372	1,270,948	934,650	336,298
Hawaii	1,187	213	183,795	133,675	50,120
Idaho	1,163	258	240,448	168,887	71,561
Illinois	11,830	2,205	1,916,172	1,368,041	548,131
Indiana	5,803	1,079	968,933	678,943	289,990
Iowa	2,842	541	499,550	344,754	154,796
Kansas	2,565	510	460,838	329,211	131,627
Kentucky	3,860	712	657,642	467,005	190,637
Louisiana	4,342	903	797,933	583,892	214,041
Maine	1,241	230	212,601	155,903	56,698
Maryland	5,042	904	790,938	580,903	210,035
Massachusetts	6,074	1,019	893,727	658,507	235,220
Michigan	9,549	1,837	1,614,784	1,170,251	444,533
Minnesota	4,610	925	821,693	581,426	240,267
Mississippi	2,697	553	505,962	366,846	139,116
Missouri	5,324	1,012	878,541	628,286	250,255
Montana	870	179	164,341	116,748	47,593
Nebraska	1,637	329	287,100	203,055	84,045
Nevada	1,530	279	250,747	185,336	65,411
New Hampshire	1,148	219	189,319	138,851	50,468
New Jersey	7,945	1,386	1,174,206	862,331	311,875
New Mexico	1,685	362	327,248	229,168	98,080
New York	18,136	3,177	2,766,208	1,949,245	816,963
North Carolina	7,195	1,285	1,156,767	847,463	309,304
North Dakota	641	129	119,288	83,419	35,869
Ohio	11,151	2,087	1,814,290	1,295,289	519,001
Oklahoma	3,278	648	609,718	442,607	167,111
Oregon	3,141	587	521,945	371,967	149,978
Pennsylvania	12,072	2,125[1]	1,765,891	1,244,103	521,788
Rhode Island	990	170	147,487	107,913	39,574
South Carolina	3,673	682	648,673	468,798	179,875
South Dakota	729	154	143,482	101,805	41,677
Tennessee	5,256	945	881,355	640,534	240,821
Texas	18,724	3,819	3,677,171	2,720,623	956,548
Utah	1,951	491	474,675	328,482	146,193
Vermont	585	110	104,533	75,590	28,943
Virginia	6,618	1,149	1,060,809	774,319	286,490
Washington	5,431	1,033	938,314	673,107	265,207
West Virginia	1,828	325	310,511	212,808	97,703
Wisconsin	5,123	1,009	860,686	601,215	259,471
Wyoming	480	104	100,369	70,185	30,184

[1] U.S. Department of Commerce, Bureau of the Census, *Current Population Reports*, Series P-25, No. 1095 at the national level and forthcoming state-level P-25 Reports.
[2] U.S. Department of Education, National Center for Education Statistics, *Common Core of Data* surveys.

Information reprinted from the *Digest of Education Statistics 1996* (NCES 96-133).

Table C.1 (continued)

School System Characteristics from Non-NAEP Sources

THE NATION'S REPORT CARD naep

	Poverty Status of 5- to 17-Year-Olds: 1994[3]		Number of Children (Birth to age 21) Served Under State-Operated Individuals With Disabilities Education Act and Chapter 1 of the Education Consolidation and Improvement Act Programs[4]	
	Number in Poverty (in thousands)	Percent in Poverty	1993-94 School Year	Percent Change: 1990-91 to 1993-94
United States	9,974	20.1	5,318,021	11.7
Alabama	157	19.5	99,760	5.1
Alaska	15	11.7	18,006	22.1
Arizona	189	23.4	69,530	21.5
Arkansas	87	20.4	53,187	11.2
California	1,550	25.3	533,807	13.7
Colorado	69	9.9	66,595	16.6
Connecticut	100	18.6	71,863	11.3
Delaware	10	9.8	15,196	6.3
District of Columbia	29	30	6,994	11.2
Florida	563	22.1	289,539	22.7
Georgia	267	18.5	123,143	20.7
Hawaii	21	12	15,248	15.8
Idaho	39	15.5	23,536	6.9
Illinois	405	18	257,986	7.9
Indiana	164	13.7	127,961	11.6
Iowa	74	13.5	63,373	4.4
Kansas	97	19.5	50,438	11.6
Kentucky	200	26.6	80,539	1.4
Louisiana	337	36.8	86,931	18
Maine	20	9.6	29,350	4.9
Maryland	143	17.2	97,998	6.6
Massachusetts	121	12.2	160,275	3.7
Michigan	326	17.9	181,251	8.6
Minnesota	115	13.7	90,918	12.4
Mississippi	138	28.2	64,153	5.3
Missouri	204	23.6	114,008	11.8
Montana	20	12.3	18,401	7
Nebraska	43	12.5	37,112	13.3
Nevada	45	16.2	25,242	36.9
New Hampshire	23	12.2	23,354	18.8
New Jersey	211	14.6	190,003	4.8
New Mexico	111	29.2	43,474	20.6
New York	769	23.5	365,697	18.9
North Carolina	206	18.4	136,513	10.9
North Dakota	15	11.6	12,440	-0.5
Ohio	448	19.5	219,875	7
Oklahoma	140	21.5	73,130	11.4
Oregon	81	13.7	63,212	14.6
Pennsylvania	400	19	210,826	-3.9
Rhode Island	24	13.3	23,582	11.9
South Carolina	121	18.7	81,930	5.4
South Dakota	32	18.2	15,907	6.1
Tennessee	206	20.1	119,146	13.6
Texas	1,084	26.8	411,917	17.5
Utah	46	9.9	51,950	8.8
Vermont	7	7	10,452	-14.8
Virginia	157	12.6	131,599	15.5
Washington	146	14.6	101,254	18.6
West Virginia	66	22	44,528	3.2
Wisconsin	120	12.1	102,412	17.8
Wyoming	12	10.7	12,480	11.4

[3] U.S. Department of Commerce, Bureau of the Census, *Decennial Census, Minority Economic Profiles,* unpublished data, and *Current Population Reports,* Series P-60, "Poverty in the United States," "Money Income of Households, Families, and Persons in the United States," and "Income, Poverty, and Valuation of Noncash Benefits," various years.

[4] U.S. Department of Education, Office of Special Education and Rehabilitative Services, *Annual Report to Congress on the Implementation of The Individuals with Disabilities Education Act,* various years, and unpublished tabulations.

Information reprinted from the *Digest of Education Statistics 1996* (NCES 96-133).

Table C.1 (continued)

School System Characteristics from Non-NAEP Sources

THE NATION'S REPORT CARD naep

	Elementary and Secondary Education Expenditures Per Capita: 1991-92[5]	Pupil-Teacher Ratios in Public Elementary and Secondary Schools: Fall 1994[6]	Estimated Average Annual Salaries of Teachers Public and Secondary Schools (current dollars)	
			NEA: 1995-96[7]	AFT: 1994-95[8]
United States	896.57	17.3	37,846	36,744
Alabama	585.31	17.2	31,307	30,545
Alaska	1,713.81	17.6	49,620	47,864
Arizona	835.69	19.3	32,484	32,223
Arkansas	705.09	17.1	29,322	28,950
California	868.44	24	42,516	40,667
Colorado	900.58	18.4	35,364	34,571
Connecticut	1,124.30	14.4	50,400	50,598
Delaware	905.69	16.6	40,533	39,076
District of Columbia	1,066.24	13.2	43,700	43,142
Florida	819.3	19.1	33,320	32,590
Georgia	805.85	16.3	34,307	32,198
Hawaii	702.2	17.9	35,807	37,443
Idaho	775.69	19.1	30,891	29,784
Illinois	801.64	17.3	41,008	39,445
Indiana	857.87	17.5	37,805	36,799
Iowa	917.11	15.7	32,376	31,511
Kansas	856.45	15.1	35,518	32,085
Kentucky	654.64	17	33,018	32,272
Louisiana	814.21	16.6	26,800	26,811
Maine	962.73	13.8	32,869	31,972
Maryland	877.49	17	41,215	40,661
Massachusetts	811.98	14.8	43,756	40,976
Michigan	1,012.79	20.1	49,168	46,575
Minnesota	1,060.85	17.5	36,937	35,948
Mississippi	639.56	17.5	27,689	26,818
Missouri	781.87	15.5	33,341	31,209
Montana	934.99	16.3	29,364	28,785
Nebraska	924.51	14.5	31,496	30,922
Nevada	897.18	18.7	36,167	38,010
New Hampshire	889.57	15.6	35,792	34,721
New Jersey	1,263.17	13.8	47,910	47,038
New Mexico	827.45	17.2	29,349	28,394
New York	1,224.39	15.2	48,115	47,612
North Carolina	788.77	16.2	30,564	30,793
North Dakota	832.42	15.3	26,969	26,317
Ohio	813.62	16.6	37,835	36,971
Oklahoma	778.17	15.5	28,909	28,745
Oregon	956.96	19.9	39,650	38,871
Pennsylvania	910.93	17.1	46,916	44,510
Rhode Island	864.33	14.7	42,160	40,729
South Carolina	800.23	16.4	31,568	30,366
South Dakota	819.08	14.4	26,346	26,037
Tennessee	586.25	18.6	33,451	31,270
Texas	885.47	15.7	32,000	31,223
Utah	830.92	24.3	30,452	28,919
Vermont	1,120.15	13.8	36,295	35,207
Virginia	854.34	14.6	34,687	33,907
Washington	1,045.76	20.2	38,025	36,160
West Virginia	865.8	14.8	32,155	31,944
Wisconsin	1,015.96	15.9	38,571	37,617
Wyoming	1,328.26	15	31,571	31,285

[5] U.S. Department of Commerce, Bureau of the Census, Government Division, *Government Finances: 1991-92*, Series GF/92-5.
[6] U.S. Department of Education, National Center for Education Statistics, *Common Core of Data* surveys.
[7] National Education Association (NEA), *Estimates of School Statistics*, and unpublished data. (Latest edition 1995-96, Copyright © 1996 by the National Education Association. All rights reserved.)
[8] American Federation of Teachers (AFT), *Survey and Analysis of Salary Trends*, various years.

Information reprinted from the *Digest of Education Statistics 1996* (NCES 96-133).

Appendix D

State-Level SD/LEP Information

This appendix contains national and state-level public school results on identification and inclusion rates for students with disabilities and LEP students. Results are presented for grade 8, the grade at which the 1996 state NAEP science assessment was conducted.

Table D.1 presents the percentages of the NAEP grade 8 public school population that were identified as students with disabilities, LEP students, or both. Those results were determined by combining data from both national and state NAEP samples. Hence, they differ slightly from those reported in Appendix A which are based on the "reporting sample" only. In the nation's public schools, 10 percent of the eighth graders were identified as students with disabilities (including those who were also identified as LEP students). The percentage with disabilities ranged from 6 percent (in Guam) to 15 percent (Florida, Massachusetts, and New Mexico), with 34 of the 44 participating jurisdictions identifying between 9 and 13 percent of eighth-graders as students with disabilities. Approximately 3 percent of the nation's eighth-graders were identified as LEP (including those who were also identified as students with disabilities). Only one jurisdiction (California) identified more than 10 percent of its population as being limited English proficient while in 29 of the 44 participating jurisdictions 2 percent or less of the eighth grade public school population was so identified.

Table D.2 presents the percentages of the NAEP grade 8 public school population in each of the state NAEP jurisdictions that were excluded from the assessment in the S1 and S2 samples. With one exception, state public results are in agreement with national results (reported in Chapter 4) in showing little evidence of any effect of revisions to the inclusion criteria on inclusion rates. In one jurisdiction, Delaware, a smaller percentage of the population was excluded in S2 using the revised criteria than in S1 using the original criteria. It should be noted, however, that because of its size, fewer schools are represented in each of the Delaware samples than in most of the other jurisdictions. Furthermore, results from the remaining jurisdictions do not suggest a clear pattern of greater inclusion for either of the sets of criteria.

In many of the jurisdictions that participated in the state assessment students with disabilities and LEP students constituted a modest percentage of the total school population. Consequently, examining exclusion rates alone may not, in some cases, provide a sufficiently sensitive measure of the effects of the inclusion criteria changes. Further analyses of national inclusion rates among students with disabilities and LEP students were included in Chapter 4.

However, due to space limitations, similar analyses at the state level were not included in the main body of the report. These analyses are included in this appendix.

Table D.3 presents the percentages of assessed students with disabilities in public schools for each of the jurisdictions participating in the state assessment. Considerable variability across jurisdictions is evident in the percentages of students with disabilities who are assessed in NAEP. The District of Columbia assessed less than 30 percent of its grade 8 students with disabilities, regardless of which inclusion criteria was used. In contrast, several jurisdictions (Alaska, Minnesota, Montana, Nebraska, North Dakota, Oregon, and Washington) assessed more than 60 percent of their students with disabilities, again regardless of which inclusion criteria were employed.

Comparisons of the S1 (sample using existing inclusion criteria) and S2 (sample using the revised inclusion criteria) inclusion percentages for students with disabilities across jurisdictions provide little evidence of a systematic or unidirectional effect due to changes in inclusion criteria. Observed inclusion percentages using the original criteria were higher for 26 of 41 jurisdictions that met sample size requirements for students with disabilities. This provides some suggestion that the revised criteria may actually have resulted in less inclusion. However, in none of these 26 jurisdictions was the difference significant. The only jurisdiction exhibiting a significant difference was Delaware, in which S2 inclusion rates were higher than those observed in S1. Averaged over jurisdictions, the S1 and S2 inclusion percentages were virtually identical (53 percent in S1 and 52 percent in S2).

Table D.4 presents LEP student inclusion percentages for the four jurisdictions participating in the NAEP state assessment in which samples of LEP students were sufficiently large to permit meaningful analysis. There were no significant differences between S1 and S2 LEP student inclusion percentages in any of these jurisdictions.

Table D.1 Percentages of Students Identified as SD*, LEP** or Both by Jurisdiction: Grade 8, Public Schools Only

	Total	SD Only	Both SD and LEP	LEP Only
Nation	**12 (0.5)**	**10 (0.5)**	**0 (0.1)**	**2 (0.3)**
Alabama	13 (0.8)	13 (0.8)	0 (0.0)	0 (0.1)
Alaska	15 (1.1)	11 (0.8)	1 (0.4)	3 (0.8)
Arizona	15 (1.3)	8 (0.6)	1 (0.3)	6 (1.2)
Arkansas	12 (0.8)	11 (0.6)	0 (0.1)	1 (0.3)
California	21 (1.2)	6 (0.4)	1 (0.2)	14 (1.1)
Colorado	12 (0.7)	9 (0.6)	0 (0.1)	3 (0.5)
Connecticut	15 (0.7)	13 (0.6)	0 (0.1)	2 (0.4)
Delaware	10 (0.7)	10 (0.7)	0 (0.1)	1 (0.1)
District of Columbia	12 (0.8)	9 (0.6)	0 (0.2)	2 (0.4)
Florida	18 (1.0)	14 (0.8)	0 (0.2)	4 (0.6)
Georgia	10 (0.5)	9 (0.5)	0 (0.1)	1 (0.3)
Hawaii	13 (0.6)	9 (0.5)	1 (0.3)	3 (0.3)
Indiana	11 (0.7)	10 (0.7)	0 (0.8)	0 (0.2)
Iowa	13 (0.8)	13 (0.8)	0 (0.1)	0 (0.2)
Kentucky	9 (0.7)	9 (0.7)	0 (0.0)	0 (0.3)
Louisiana	11 (0.7)	10 (0.7)	0 (0.1)	1 (0.3)
Maine	13 (0.7)	12 (0.6)	0 (0.1)	1 (0.3)
Maryland	12 (0.7)	11 (0.7)	0 (0.2)	1 (0.3)
Massachusetts	17 (1.0)	14 (0.8)	0 (0.1)	2 (0.5)
Michigan	10 (0.7)	8 (0.7)	0 (0.1)	1 (0.3)
Minnesota	11 (0.8)	10 (0.8)	0 (0.2)	1 (0.3)
Mississippi	11 (0.7)	10 (0.7)	0 (0.1)	0 (0.0)
Missouri	13 (0.7)	12 (0.6)	0 (0.1)	1 (0.3)
Montana	9 (0.5)	9 (0.5)	0 (0.1)	0 (0.1)
Nebraska	11 (0.6)	10 (0.6)	0 (0.1)	1 (0.2)
New Mexico	20 (0.7)	13 (0.7)	2 (0.3)	5 (0.6)
New York	15 (1.0)	9 (0.7)	0 (0.2)	5 (0.8)
North Carolina	10 (0.5)	9 (0.5)	0 (0.0)	1 (0.2)
North Dakota	9 (0.5)	9 (0.5)	0 (0.1)	0 (0.1)
Oregon	12 (0.9)	10 (0.8)	0 (0.2)	2 (0.6)
Rhode Island	17 (0.7)	12 (0.7)	0 (0.1)	4 (0.3)
South Carolina	10 (0.7)	10 (0.7)	0 (0.0)	0 (0.1)
Tennessee	12 (1.0)	12 (0.9)	0 (0.1)	0 (0.3)
Texas	17 (1.3)	11 (0.8)	1 (0.2)	6 (1.1)
Utah	9 (0.6)	8 (0.6)	0 (0.1)	1 (0.2)
Vermont	14 (0.9)	13 (0.9)	0 (0.1)	1 (0.2)
Virginia	12 (0.7)	10 (0.6)	0 (0.2)	2 (0.4)
Washington	11 (0.8)	8 (0.5)	0 (0.1)	3 (0.7)
West Virginia	12 (0.8)	12 (0.8)	0 (0.0)	0 (0.1)
Wisconsin	11 (0.8)	10 (0.8)	0 (0.1)	1 (0.2)
Wyoming	10 (0.6)	9 (0.6)	0 (0.1)	1 (0.2)
DDESS	10 (1.2)	8 (1.1)	0 (0.0)	1 (0.4)
DoDDS	8 (0.6)	6 (0.5)	0 (0.1)	2 (0.3)
Guam	9 (0.8)	6 (0.6)	0 (0.2)	3 (0.6)

The standard errors of the estimated percentages appear in parentheses.

*Students with Disabilities

**Limited English Proficient Students

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools

DoDDS: Department of Defense Dependents Schools (Overseas)

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table D.2 Percentages of Students Excluded From the Assessment by Jurisdiction: Grade 8, Public Schools Only

	S1: Using Original Inclusion Criteria	S2: Using Revised Inclusion Criteria
Alabama	7 (0.7)	9 (1.4)
Alaska	3 (0.9)	5 (0.9)
Arizona	7 (1.1)	6 (0.7)
Arkansas	8 (1.1)	7 (1.2)
California	10 (1.3)	9 (0.8)
Colorado	5 (0.6)	7 (0.8)
Connecticut	8 (0.7)	9 (0.8)
Delaware	7 (0.7)	2 (0.4) *
District of Columbia	8 (1.4)	9 (1.6)
Florida	10 (1.1)	10 (1.0)
Georgia	7 (0.7)	5 (0.7)
Hawaii	7 (0.9)	6 (0.8)
Indiana	6 (0.8)	6 (0.6)
Iowa	6 (1.0)	5 (0.9)
Kentucky	4 (0.6)	4 (0.8)
Louisiana	7 (0.9)	6 (0.6)
Maine	6 (0.7)	7 (0.7)
Maryland	5 (0.9)	5 (0.9)
Massachusetts	6 (1.0)	7 (0.9)
Michigan	5 (0.9)	5 (0.8)
Minnesota	4 (0.8)	4 (0.6)
Mississippi	7 (0.7)	6 (0.7)
Missouri	6 (0.9)	6 (0.8)
Montana	3 (0.4)	4 (0.6)
Nebraska	4 (0.6)	4 (0.5)
New Mexico	10 (0.8)	9 (0.9)
New York	6 (0.9)	10 (1.2)
North Carolina	4 (0.5)	5 (0.5)
North Dakota	2 (0.4)	2 (0.3)
Oregon	4 (0.6)	5 (0.9)
Rhode Island	7 (1.0)	7 (0.8)
South Carolina	6 (0.8)	6 (0.8)
Tennessee	5 (0.9)	4 (0.7)
Texas	9 (1.0)	8 (0.9)
Utah	5 (0.6)	4 (0.6)
Vermont	5 (0.8)	7 (0.9)
Virginia	7 (0.7)	6 (0.8)
Washington	4 (0.7)	4 (0.6)
West Virginia	8 (0.7)	7 (0.7)
Wisconsin	7 (1.1)	7 (0.8)
Wyoming	3 (0.6)	4 (0.5)
DDESS	3 (0.9)	8 (2.7)
DoDDS	3 (0.7)	3 (0.6)

The standard errors of the estimated percentages appear in parentheses.

* Indicates a significant difference between S1 and S2 results.

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools

DoDDS: Department of Defense Dependents Schools (Overseas)

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table D.3 Percentages of Students with Disabilities Included in the Assessment by Jurisdiction: Grade 8, Public Schools Only

THE NATION'S REPORT CARD naep

	S1: Using Original Inclusion Criteria	S2: Using Revised Inclusion Criteria
Alabama	45 (4.7)	39 (7.1)
Alaska	73 (6.8)	65 (5.7)
Arizona	39 (7.2)	45 (5.1)
Arkansas	30 (5.3)	39 (7.6)
California	50 (4.5)	48 (5.6)
Colorado	57 (4.9)	50 (6.0)
Connecticut	51 (3.9)	41 (4.0)
Delaware	39 (9.3)	80 (4.2) *
District of Columbia	28 (5.0)	27 (6.6)
Florida	48 (5.3)	44 (5.1)
Georgia	36 (6.0)	53 (6.7)
Hawaii	45 (8.3)	56 (6.9)
Indiana	47 (5.7)	43 (3.7)
Iowa	59 (5.2)	64 (5.5)
Kentucky	65 (5.4)	53 (8.0)
Louisiana	43 (7.5)	42 (5.7)
Maine	59 (3.3)	46 (5.7)
Maryland	58 (6.1)	55 (6.3)
Massachusetts	70 (5.4)	59 (4.6)
Michigan	44 (8.0)	44 (6.3)
Minnesota	62 (6.6)	68 (4.9)
Mississippi	34 (4.9)	45 (5.7)
Missouri	53 (5.9)	52 (5.0)
Montana	65 (4.4)	64 (5.2)
Nebraska	67 (3.9)	66 (4.4)
New Mexico	54 (5.1)	52 (7.3)
New York	51 (5.4)	49 (9.0)
North Carolina	58 (5.1)	49 (5.1)
North Dakota	72 (4.4)	78 (4.0)
Oregon	68 (4.7)	62 (6.2)
Rhode Island	58 (5.8)	63 (4.1)
South Carolina	43 (5.8)	35 (4.6)
Tennessee	56 (5.7)	70 (4.8)
Texas	53 (5.3)	47 (7.0)
Utah	43 (5.8)	53 (4.3)
Vermont	64 (5.5)	58 (5.3)
Virginia	43 (4.6)	37 (5.9)
Washington	64 (6.3)	67 (5.7)
West Virginia	34 (4.0)	42 (5.1)
Wisconsin	37 (6.4)	36 (5.8)
Wyoming	66 (3.6)	59 (6.0)

The standard errors of the estimated percentages appear in parentheses.

* Indicates a significant difference between S1 and S2 results.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table D.4 Percentage of Limited English Proficient Students Included in the Assessment: Grade 8 Public School Only

	S1: Using Original Inclusion Criteria	S2: Using Revised Inclusion Criteria
Arizona	57 (9.8)	70 (7.6)
California	54 (6.6)	61 (3.6)
New Mexico	28 (12.8)	57 (8.1)
Texas	45 (8.1)	61 (7.0)

The standard errors of the estimated percentages appear in parentheses.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Appendix E

Standard Errors

The comparisons presented in this report are based on statistical tests that consider the magnitude of the difference between group averages or percentages and the standard errors of those statistics. The following appendix contains the standard errors for the averages and percentages discussed in Chapters 2 and 4.

Table E.1 — Scale Score Standard Errors — Grade 4

THE NATION'S REPORT CARD naep

	Percentage of Students	Average Scale Score	Selected Percentiles 10th	25th	50th	75th	90th
All Students		**0.8**	**2.0**	**1.0**	**1.0**	**0.7**	**1.2**
Region							
Northeast	1.5	1.8	3.8	3.2	1.6	1.6	1.3
Southeast	1.7	2.0	1.9	3.3	2.2	2.4	1.4
Central	1.3	2.1	4.4	3.7	1.6	1.7	1.8
West	1.9	2.0	3.1	1.8	2.6	2.4	2.0
Gender							
Male	0.6	0.9	1.9	1.3	1.0	0.8	1.1
Female	0.6	0.9	2.2	1.1	0.9	0.9	1.2
Race/Ethnicity							
White	0.5	0.9	0.9	1.3	1.0	0.9	1.1
Black	0.2	1.9	2.7	2.3	1.9	2.0	4.0
Hispanic	0.5	1.7	3.5	2.3	2.1	1.5	1.2
Asian/Pacific Islander	0.2	3.6	6.4	7.7	4.0	4.6	6.8
American Indian	0.2	3.8	3.2	9.0	5.0	7.6	7.6
Parents' Highest Education Level							
Did Not Finish High School	0.3	2.2	3.0	2.6	2.8	3.0	4.7
Graduated From High School	0.7	1.5	2.9	2.6	2.5	1.1	1.6
Some Education After High School	0.4	1.6	3.5	3.3	2.7	1.5	2.2
Graduated From College	1.4	1.0	1.8	2.2	0.8	0.7	1.2
I Don't Know	0.9	1.1	2.4	1.4	1.1	1.7	2.0
Type of School							
Public Schools	1.7	0.9	1.3	1.8	1.2	0.9	1.4
Nonpublic Schools:	1.7	1.8	3.0	2.5	1.6	1.5	1.7
Catholic	1.4	1.9	3.7	2.1	2.6	1.3	1.9
Other Private Schools	1.2	3.6 !	7.9 !	2.8 !	2.2 !	4.0 !	4.8 !
Title I Participation							
Participated	1.7	1.9	2.5	2.0	2.9	1.8	3.3
Did Not Participate	1.7	1.0	2.3	1.6	1.1	1.1	1.3
Free/Reduced-Price Lunch Eligibility							
Eligible	1.8	1.3	1.7	1.9	1.5	1.4	1.6
Not Eligible	2.2	0.9	1.8	1.4	1.2	1.1	0.9
Information Not Available	1.9	3.5	5.9	5.0	4.3	2.3	2.9

Scale scores range from 0 to 300.

NAEP science scales were developed independently for each grade assessed; therefore, results are not comparable across grades.

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (see Appendix A).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table E.2 Scale Score Standard Errors — Grade 8

	Percentage of Students	Average Scale Score	Selected Percentiles 10th	25th	50th	75th	90th
All Students		**0.9**	**1.0**	**1.1**	**0.8**	**1.4**	**1.4**
Region							
Northeast	1.6	2.6	3.7	2.4	4.5	2.2	4.3
Southeast	2.2	1.9	2.4	3.1	2.2	1.8	1.1
Central	0.5	2.5	3.3	4.2	3.5	2.0	2.1
West	2.4	2.2	2.9	2.7	2.6	2.3	2.8
Gender							
Male	1.0	1.0	1.8	1.6	1.0	1.4	1.6
Female	1.0	1.1	1.0	1.4	1.3	1.5	2.3
Race/Ethnicity							
White	0.2	1.1	1.4	1.1	1.3	0.9	1.5
Black	0.1	1.1	1.4	1.9	1.2	2.8	2.0
Hispanic	0.2	1.7	2.6	1.9	1.4	2.8	1.9
Asian/Pacific Islander	0.3	3.1	4.4	2.7	3.5	3.6	5.0
American Indian	0.2	4.1	7.0	5.7	5.8	4.4	4.6
Parents' Highest Education Level							
Did Not Finish High School	0.4	1.9	2.9	2.6	4.0	5.4	3.9
Graduated From HS	0.9	1.5	1.9	2.5	1.6	1.8	1.2
Some Education After HS	0.7	1.1	1.5	1.3	2.9	2.0	1.5
Graduated From College	1.2	1.2	1.5	1.0	1.2	1.6	2.0
I Don't Know	0.5	2.4	5.5	3.6	3.9	3.3	6.5
Type of School							
Public Schools	1.4	0.9	1.6	1.3	0.9	1.1	1.3
Nonpublic Schools:	1.4	2.5	8.1	3.1	3.0	2.8	2.1
Catholic	0.9	2.7	4.6	3.5	3.6	2.7	2.7
Other Private Schools	1.0	4.2 !	6.6 !	5.0 !	5.1 !	5.2 !	6.0 !
Title I Participation							
Participated	2.1	4.6	4.8	4.5	5.5	6.1	5.9
Did Not Participate	2.1	1.1	2.3	1.6	1.2	1.1	1.5
Free/Reduced-Price Lunch Eligibility							
Eligible	1.5	1.6	2.9	2.0	2.0	1.7	2.6
Not Eligible	3.3	1.2	2.2	1.1	1.6	0.9	2.0
Information Not Available	4.1	2.9	4.3	3.7	3.2	2.4	3.1

Scale scores range from 0 to 300.

NAEP science scales were developed independently for each grade assessed; therefore, results are not comparable across grades.

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (see Appendix A).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table E.3 Scale Score Standard Errors — Grade 12

THE NATION'S REPORT CARD naep

	Percentage of Students	Average Scale Score	Selected Percentiles 10th	25th	50th	75th	90th
All Students		**0.9**	**1.0**	**1.2**	**1.0**	**1.2**	**1.2**
Region							
Northeast	1.3	2.8	2.8	4.7	3.3	3.8	1.4
Southeast	1.8	1.4	2.3	2.2	1.4	1.4	1.4
Central	0.9	2.0	4.1	3.9	1.8	2.4	2.2
West	1.8	2.3	2.6	4.3	3.2	2.1	2.3
Gender							
Male	0.9	1.2	1.7	1.3	1.3	1.9	1.9
Female	0.9	0.9	1.5	1.6	1.5	0.9	1.0
Race/Ethnicity							
White	0.4	1.0	2.2	0.9	1.3	1.4	0.9
Black	0.4	1.5	2.1	1.3	2.7	2.0	1.8
Hispanic	0.3	2.3	3.1	4.1	2.1	2.5	1.4
Asian/Pacific Islander	0.2	2.9	5.6	5.4	3.1	2.6	2.2
American Indian	0.2	4.7 !	4.9 !	6.0 !	5.8 !	13.2 !	9.8 !
Parents' Highest Education Level							
Did Not Finish High School	0.5	1.8	5.8	2.3	3.8	1.5	4.5
Graduated From HS	0.8	1.5	2.2	1.2	1.9	1.6	5.3
Some Education After HS	0.7	1.1	1.2	2.0	1.2	1.5	2.2
Graduated From College	1.4	1.0	2.7	0.8	1.2	1.0	1.0
I Don't Know	0.3	3.1	7.0	4.8	3.4	5.3	7.5
Type of School							
Public Schools	1.7	1.0	1.7	1.3	1.1	1.4	1.2
Nonpublic Schools:	1.7	2.2	4.5	2.2	2.5	2.2	2.6
Catholic	1.3	2.5	8.8	3.1	1.2	3.6	2.7
Other Private Schools	1.2	3.9 !	8.9 !	5.3 !	4.7 !	3.9 !	4.6 !
Title I Participation							
Participated	1.1	5.7 !	3.6 !	4.7 !	7.4 !	8.1 !	10.8 !
Did Not Participate	1.1	0.9	1.5	1.4	0.9	1.0	1.0
Free/Reduced-Price Lunch Eligibility							
Eligible	1.2	1.9	1.9	2.6	2.9	3.7	3.3
Not Eligible	3.9	0.9	1.5	1.2	1.2	1.1	1.2
Information Not Available	4.0	2.9	4.0	3.6	3.1	2.5	2.7

Scale scores range from 0 to 300.

NAEP science scales were developed independently for each grade assessed; therefore, results are not comparable across grades.

! Statistical tests involving this value should be interpreted with caution. Standard error estimates may not be accurately determined and/or the sampling distribution of the statistic does not match statistical test assumptions (see Appendix A).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table E.4

Scale Score Standard Errors by Jurisdiction for Grade 8 Public Schools

	MEAN	10th	25th	50th	75th	90th
Maine	1.0	1.6	1.6	1.2	1.0	1.0
North Dakota	0.8	1.3	1.5	0.7	1.1	1.0
Montana ‡	1.2	2.6	1.7	1.2	0.6	1.9
Wisconsin ‡	1.7	2.8	1.9	1.6	1.0	1.0
Minnesota	1.3	3.1	1.6	1.4	1.2	2.9
Iowa ‡	1.2	1.2	1.5	1.0	1.0	0.7
Wyoming	0.6	1.3	0.8	0.8	0.8	0.8
Nebraska	1.0	1.5	1.2	1.0	1.2	1.3
Vermont ‡	1.0	2.2	1.5	1.2	1.4	2.3
Massachusetts	1.4	2.6	2.5	2.0	1.0	1.3
Utah	0.8	2.0	1.3	0.7	1.1	1.2
Connecticut	1.3	2.5	1.8	1.5	1.2	1.1
DoDDS	0.7	0.9	1.4	0.9	1.0	0.9
Oregon	1.6	3.3	2.3	1.4	1.2	1.4
Colorado	0.9	1.9	1.2	1.0	1.4	0.9
Michigan ‡	1.4	2.0	2.3	1.3	1.2	1.6
Indiana	1.4	2.0	1.6	1.2	1.7	1.9
DDESS	1.1	2.4	1.5	1.6	1.9	2.1
Alaska ‡	1.3	2.9	1.9	1.4	1.6	1.7
Missouri	1.2	2.7	1.8	1.1	1.2	1.6
Washington	1.3	2.3	2.2	1.5	1.6	1.4
Virginia	1.6	3.1	2.6	2.8	1.8	2.1
Rhode Island	0.8	1.4	0.9	1.0	1.2	1.1
Nation	**0.9**	**1.6**	**1.3**	**0.9**	**1.1**	**1.3**
Kentucky	1.2	2.7	1.6	1.3	1.1	1.2
West Virginia	0.9	2.1	1.5	1.2	1.0	1.1
North Carolina	1.2	1.9	1.5	1.2	1.7	1.0
New York ‡	1.6	2.9	2.5	2.0	1.4	1.3
Maryland ‡	1.5	3.4	1.5	2.2	2.0	1.9
Texas	1.8	2.5	2.6	1.7	1.7	1.6
Arizona	1.6	2.9	2.3	1.7	1.7	2.5
Arkansas ‡	1.3	2.6	1.7	1.8	1.4	2.1
Tennessee	1.8	3.4	2.3	1.8	1.2	2.1
Florida	1.6	2.5	2.8	1.6	1.6	1.7
Georgia	1.4	2.8	1.4	1.4	1.9	1.7
Delaware	0.8	2.9	1.8	0.9	0.8	0.9
New Mexico	1.0	1.0	1.7	1.3	1.3	1.7
Alabama	1.6	2.3	1.5	1.8	1.8	1.8
South Carolina ‡	1.5	2.2	1.9	1.8	1.9	1.9
California	1.7	1.6	3.3	1.9	1.8	1.4
Hawaii	0.7	1.9	1.5	0.9	0.9	1.5
Mississippi	1.4	3.0	1.6	1.5	1.3	1.5
Lousiana	1.6	2.4	2.1	1.4	2.2	2.7
Guam	1.1	2.2	1.9	1.8	2.2	1.8
District of Columbia	0.7	3.2	1.8	1.0	1.4	1.7

Scale scores range from 0 to 300.

‡ Indicates that the jurisdiction did not satisfy one or more of the guidelines for public school participation rates (see Appendix A).

DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools

DoDDS: Department of Defense Dependents Schools (Overseas)

National results are based on the national assessment sample of public schools, not on aggregated state assessment program samples (see Appendix A).

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table E.5

Standard Errors for the Percentage of National Population Identified as SD*, LEP**, or Both: Public and Nonpublic Schools Combined

	Total	SD* Only	LEP** Only	Both SD* and LEP**
Grade 4	0.8	0.6	0.5	0.2
Grade 8	0.5	0.4	0.2	0.1
Grade 12	0.4	0.3	0.3	0.1

* Students with Disabilities.

** Limited English Proficient Students.

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table E.6

Standard Errors for the Percentage of National Population Excluded From the Assessment: Public and Nonpublic Schools Combined

	S2: Using Revised Inclusion Criteria	S3: Using Revised Criteria and Providing Accommodations
Grade 4	0.9	0.7
Grade 8	0.4	0.4
Grade 12	0.4	0.3

SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Table E.7

Standard Errors for the Percentage of Students with Disabilities and Limited English Proficient Students in the National Population Included in the Assessment: Public and Nonpublic Schools Combined

	SD*		LEP**	
	S2: Using Revised Inclusion Criteria	S3: Using Revised Criteria And Providing Accommodations	S2: Using Revised Inclusion Criteria	S3: Using Revised Criteria And Providing Accommodations
Grade 4				
Assessed Under Standard Conditions	3.9	3.5	7.3	7.0
Assessed With Accommodations		5.0		3.4
Total Assessed	3.9	4.5	7.3	7.6
Grade 8				
Assessed Under Standard Conditions	3.4	3.6	6.0	6.0
Assessed With Accommodations		3.6		3.2
Total Assessed	3.4	3.2	6.0	7.0
Grade 12				
Assessed Under Standard Conditions	4.4	4.7	9.3	4.4
Assessed With Accommodations		2.8		1.2
Total Assessed	4.4	3.9	9.3	4.1

* Students with Disabilities.
** Limited English Proficient Students.
SOURCE: National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Science Assessment.

Acknowledgments

This report is the culmination of the effort of many individuals who contributed their considerable knowledge, experience, and creativity to the NAEP 1996 science assessment. The NAEP 1996 science assessment was a collaborative effort among staff from the National Center for Education Statistics (NCES), the National Assessment Governing Board (NAGB), Educational Testing Service (ETS), Westat, and National Computer Systems (NCS). In addition, the program benefited from the contributions of hundreds of individuals at the state and local levels — governors, chief state school officers, state and district test directors, state coordinators, and district administrators — who tirelessly provided their wisdom, experience, and hard work. Most importantly, NAEP is grateful to the thousands of students and hundreds of teachers and administrators who made the assessment possible.

The NAEP 1996 science assessment was funded through NCES, in the Office of Educational Research and Improvement of the U.S. Department of Education. The Commissioner of Education Statistics, Pascal D. Forgione, and the NCES staff — Sue Ahmed, Peggy Carr, Arnold Goldstein, Steven Gorman, Larry Ogle, Gary W. Phillips, Sharif Shakrani, Maureen Treacy, and Alan Vanneman — worked closely and collegially with the authors to produce this report. The authors were also provided invaluable advice and guidance by the members of the National Assessment Governing Board and NAGB staff. In particular, the authors are indebted to Arnold Goldstein of NCES for his daily efforts to coordinate the activities of the many people who contributed to this report.

The NAEP project at ETS is housed in the Center for the Assessment of Educational Progress under the direction of Paul Williams. The NAEP 1996 assessments were directed by Stephen Lazer and John Mazzeo. Tom Corley, Lee Jones, Tim Ligget, Beth Nichols, Christine O'Sullivan, Amy Pearlmutter, Will Pfeiffenberger, Mario Yepes-Baraya, and Ann Marie Zolandz worked with the Science Instrument Development committee to develop the assessment instrument. Sampling and data collection activities were conducted by Westat under the direction of Rene Slobasky, Nancy Caldwell, Keith Rust, and Dianne Walsh. Printing, distribution, scoring, and processing activities were conducted by NCS under the direction of Brad Thayer, Patrick Bourgeacq, Charles Brungardt, Mathilde Kennel, Linda Reynolds, and Brent Studer.

The statistical and psychometric activities for NAEP at ETS are directed by Nancy Allen, John Barone, James Carlson, and Juliet Shaffer. The analyses presented in this report were led by John Donoghue and Steven Isham with assistance from Jinming Zhang, Spencer Swinton, Lois Worthington, Inge Novatkoski, Bruce Kaplan, Dave Freund, and Kate Pashley.

The design and production of the report was overseen by Carol Errickson, Karen Miller, and Clyde Reese. The considerable production efforts were completed by Loretta Casalaina, Kelly Gibson (cover design), Sharon Davis-Johnson, Alice Kass and Barbette Tardugno. Editorial assistance was provided by Lynn Jenkins. The production of the World Wide Web version of this report was led by Patricia O'Reilly with assistance from Jim Rura and Debbie Kline.

Many thanks are due to the numerous reviewers, both internal and external to NCES and ETS. The comments and critical feedback of the following reviewers are reflected in this report: Mary Lyn Bourque, Audrey Champagne, Lawrence Feinberg, Henry Heikkinen, Andrew Kolstad, Michelle Leon, Laura Lippman, Marilyn McMillen, Wayne Martin, Senta Raizen, Thomas Sachse, Sylvia Ware, Arthur Williams, David Williams, and Shi-Chang Wu.